# Journal Pre-proof

A free energy principle for generic quantum systems

Chris Fields, Karl Friston, James F. Glazebrook, Michael Levin

Please cite this article as: Fields, C., Friston, K., Glazebrook, J.F., Levin, M., A free energy principle for generic quantum systems, *Progress in Biophysics and Molecular Biology* (2022), doi: https://doi.org/10.1016/j.pbiomolbio.2022.05.006.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

# A free energy principle for generic quantum systems

Chris Fields[a]*, Karl Friston[b], James F. Glazebrook[c],[d] and Michael Levin[e]

[a] *23 Rue des Lavandières, 11160 Caunes Minervois, FRANCE*
[b] *Wellcome Centre for Human Neuroimaging, University College London, London, WC1N 3AR, UK*
[c] *Department of Mathematics and Computer Science, Eastern Illinois University, Charleston, IL 61920 USA*
[d] *Adjunct Faculty, Department of Mathematics, University of Illinois at Urbana-Champaign, Urbana, IL 61801 USA*
[e] *Allen Discovery Center at Tufts University, Medford, MA 02155 USA*

May 19, 2022

**Abstract**

The Free Energy Principle (FEP) states that under suitable conditions of weak coupling, random dynamical systems with sufficient degrees of freedom will behave so as to minimize an upper bound, formalized as a variational free energy, on surprisal (a.k.a., self-information). This upper bound can be read as a Bayesian prediction error. Equivalently, its negative is a lower bound on Bayesian model evidence (a.k.a., marginal likelihood). In short, certain random dynamical systems evince a kind of self-evidencing. Here, we reformulate the FEP in the formal setting of spacetime-background free, scale-free quantum information theory. We show how generic quantum systems can be regarded as observers, which with the standard freedom of choice assumption become agents capable of assigning semantics to observational outcomes. We show how such agents minimize Bayesian prediction error in environments characterized by uncertainty, insufficient learning, and quantum contextuality. We show that in its quantum-theoretic formulation, the FEP is asymptotically equivalent to the Principle of Unitarity. Based on these results, we suggest that biological systems employ quantum coherence as a computational resource and – implicitly – as a communication resource. We summarize a number of problems for future research,

*Corresponding author at: 23 Rue des Lavandières, 11160 Caunes Minervois, FRANCE; *E-mail address*: fieldsres@gmail.com

particularly involving the resources required for classical communication and for detecting and responding to quantum context switches.

# Contents

# 1    Introduction

Since its introduction as a theory of brain function [1, 2, 3, 4], the variational Free Energy Principle (FEP) has been extended into an explanatory framework for living systems at all scales [5, 6, 7, 8, 9], and shown to characterize, in its most general form, all random dynamical systems that remain measurable, and hence identifiable as persistent and separable entities, over macroscopic times [10]. To summarize, it is shown in [10] that any system that has a non-equilibrium steady state (NESS) solution to its density dynamics i) possesses an internal dynamics that is conditionally independent of the dynamics of its environment, and ii) will continuously "self-evidence" by returning its state to (the vicinity of) its NESS. The FEP is the statement that any measurable, i.e. bounded and macroscopically persistent, system will behave so as to satisfy these requirements. In particular, the classical FEP as developed in [10] considers the joint environment-agent system as a random dynamical system that possesses an attracting set. By placing particular constraints on the coupling among systemic states (e.g., with sparse coupling in flow operators or stochastic differential equations), one can partition the joint state space into external, internal and blanket (MB; see [11] for an informal review) states. In turn, the blanket states are partitioned into sensory states that mediate the influence of external states on internal states and active states that mediate the influence of internal states on external states. Crucially, this particular partition imposes conditional independence between internal and external states, given the blanket states. The final move in the classical FEP is to induce a variational density $Q_\mu(\eta)$ over external states $\eta$ that is parameterised by internal states $\mu$. It is then fairly straightforward to show that the expected flow of internal (and active) states can be expressed as a gradient flow on a variational free energy. This free energy is effectively the divergence between the variational density encoded by internal states and the density over external states conditioned on the blanket states. This licences an interpretation of internal and active states in terms of active inference, or a Bayesian mechanics, in which their expected flow can be read as perception and action, respectively. In other words, active inference is a process of Bayesian belief updating that incorporates active exploration of the environment. In essence, this Bayesian belief updating is one way of reading a generalised synchrony between two random dynamical systems that are coupled vicariously via a Markov blanket. In what follows, we revisit this classical perspective and ask whether the implicit statistical entanglement has a more fundamental basis in quantum mechanics.

Self-organization, the FEP tells us, is not a rare, special case, but a ubiquitous feature of physical systems with sufficient dynamical stability to be called "things." All "things," in particular, self-organize MBs comprising "sensory" states that encode incoming information and thus mediate the influence of external states on internal states, and "active" states that encode outgoing information and thus mediate the influence of internal states on external states. This partitioning of "things" into internal and MB states means that every "thing" can be construed as a certain kind of "particle" – a particle that is in open exchange with external states via its MB. In short, the MB of any such "particle" underwrites conditional independence between its internal states and the external states of its environment by localizing and thereby restricting information exchange; hence, the MB can be viewed as

separating internal from external states, while mediating their exchange.

As noted in [10], generalizing the FEP to characterize the behavior of all "things" substantially weakens the traditional distinction between "cognition" and merely "physical" dynamics, and hence weakens the even deeper, pretheoretical [12] distinction between "agents" and mere "objects" [13]. Treating physical interaction as information exchange – "observation" of the environment followed by "action" upon it – redescribes the "mechanical" process of returning to the NESS – as a random global (a.k.a., pullback) attractor – in terms of inference. This can be read as "active inference" [4, 5, 6, 10] in which the existential integrity of the MB, and hence of the "self" – "world" distinction [14], can be maintained in the face of environmental fluctuations by changing internal states (via sensory states: c.f., perception) or changing external states (via active states: c.f., action). It refocuses the discussion, in other words, from abstract trajectories (or flows) in state space to the MB itself as a concrete locus of "identity" in the form of persistent measurability, and hence of "self-evidencing" via active inference to maintain that identity.

It is worth emphasizing that the FEP is a theory of what can be termed, in the language of cognitive or computer science [15], the *functional architectures* of physical systems. It describes the information processing functions that physical systems execute, but is agnostic about their implementation. As first established by Church [16] and Turing [17], information processing functions are implementation-independent in principle; provided the generic capabilities of a Turing machine (or of AND and NOT gates), choice of implementation determines resource requirements including time, space and free energy, but does not affect what can be computed. Implementation independence – also sometimes termed "multiple realizability" [18, 19] – underlies the black box methodology of classical cybernetics [20], in particular Moore's theorem [21] showing that the internal processes of a black box cannot be determined by finite observation. Indeed an MB can be viewed as implementing the input/output (I/O) interface of a black box; in more general terms, it can be viewed as implementing the application programming interface (API) of a modularized computing system [22]. An MB can, therefore, be viewed as encapsulating, in the sense of [15] or of computer science generally, information about implementation. In the language of [23], an MB implements a "quantum protectorate" of the system it encloses.

The idea that all physical systems, including the environment at large, can be considered "observers" that also act on their surroundings to "prepare" them for subsequent observations has become commonplace in quantum theory, largely replacing the "wave-function collapse" postulate of traditional quantum mechanics [24] with interaction-induced decoherence (i.e., dissipation of quantum coherence) as the generator of classical information [25, 26, 27, 28, 29].[1] Indeed while quantum theory was originally developed – and is still widely regarded – as a theory specifically applicable at the atomic scale and below, since

---

[1]Variants of quantum theory that postulate a physical collapse mechanism also invoke interaction, e.g. with gravity or a "noise" field, to generate classicality; see [30] for review. We will not discuss here the question of "interpretations" of quantum theory; see [31] for a thorough review and [32] for a more recent taxonomy highlighting fundamental assumptions. Both decoherence and entanglement, in particular, have different physical meanings in different interpretations, although their mathematical representations and observable effects are interpretation-independent. Our general approach can be viewed as replacing the

the pioneering work of Wheeler [33], Feynman [34], and Deutsch [35], it has, over the past few decades, been reformulated as a scale-free information theory [36, 37, 38, 39, 40, 41] and is increasingly viewed as a theory of the process of observation itself [42, 43, 44, 45, 46, 47]. This newer understanding of quantum theory fits comfortably with the generalization of the FEP, and hence of self-evidencing and active inference, to all "things" as outlined in [10], and with the general view of observation under uncertainty as inference.

In what follows, we take the natural next step from [10], formulating the FEP as a generic principle of quantum information theory. We show, in particular, that the FEP emerges naturally in any setting in which an "agent" or "particle" deploys quantum reference frames (QRFs), namely, physical systems that give observational outcomes an operational semantics [48, 49], to identify and characterize the states of other systems in its environment. This reformulation removes two central assumptions of the formulation in terms of random dynamical systems employed in [10]: the assumption of a spacetime embedding (or "background" in quantum-theoretic language) and the assumption of "objective" or observer-independent randomness. It further reveals a deep relationship between the ideas of local ergodicity and system identifiability, and hence the idea of "thingness" highlighted in [10], and the quantum-theoretic idea of separability, i.e., the absence of quantum entanglement, between physical systems.

Any quantum system that can be distinguished from its environment over time can, therefore, be regarded as self-organizing and self-evidencing as described in [10]. We then show that when the FEP is taken to an asymptotic limit, it drives systems away from separability towards entanglement, and hence towards a supraclassical statistical coupling between each "thing" and its environment – between the observer and the observed. In this, the FEP reproduces the Principle of Unitary, i.e. the Principle of Conservation of Information, which similarly drives all interacting systems asymptotically toward entanglement. Hence the FEP is, in an important sense, an alternative statement of the Principle of Unitarity, the most fundamental principle of quantum theory. It therefore applies to a much broader array of systems than would fall under an intuitive idea of "thingness," e.g. to quantum fields, and applies in principle from the Planck scale to cosmological scales. Formulating the FEP as a generic principle of quantum information theory thus substantially expands the range of systems to which "cognitive" or information-processing concepts reasonably apply.

We begin by reviewing in §2 the basic principles of quantum theory from an information-theoretic perspective, limiting the formalism to focus on the physical meaning of the theory. Using the category-theoretic [50, 51] formalism of Channel Theory – developed by Barwise and Seligman [52] to formalize the operational semantics of natural languages – we develop a generic formal representation of QRFs and show how the noncommutativity of QRFs induces quantum contextuality [53, 54], a nonclassical effect demonstrating the presence of entanglement between distinct physical degrees of freedom. We develop in §3 a generic, formal description of how one quantum system identifies another quantum system as a

---

"measurement problem" – that such interpretations are designed to solve – with an explicit, quantum-informational theory of measurement. This theory is, we show, just the theory of the FEP.

persistent entity – a "thing" – and measures, records, and compares its states by deploying specific sequences of QRFs. This identification and measurement process depends critically on breaking thermodynamic symmetries, and therefore on system-specific flows of energy. These sections together provide a representation of generic quantum systems as observers, or in the language of Gell-Mann and Hartle [55] "information gathering and using systems" (IGUSs), that is free of scale and spacetime embedding (i.e. "background") dependent assumptions. It also treats all probabilities as observer-relative. We then show in §4 how the FEP emerges in this setting and analyze its asymptotic behavior; in particular, we consider how the FEP addresses the fundamental problem posed by quantum context switches. We conclude in §5 by discussing the relevance of these results to a scale-independent understanding of biological systems as "particles" that interact with other "particles," whether these are other organisms, "objects," or an undifferentiated environment. We consider in particular the circumstances in which this "particle" nature can break down, and suggest that well-designed experiments may be expected to detect quantum context switches or violations of the Bell [56] or Leggett-Garg [57] inequalities, any of which indicate entanglement, by macroscopic biological systems under ordinary conditions.

# 2 Physical interaction as information exchange

## 2.1 What is "quantum"?

When physical interaction is viewed as information exchange, why it is "quantum" becomes obvious: the fundamental quantum of information is one bit, one unit of entropy, that one system exchanges with another. One bit, one quantum of information, is the answer to one yes/no question. Planck's quantum of action $\hbar$ is then naturally regarded as the action (energy · time) required to obtain one bit via any physical interaction. The energy required to irreversibly obtain one bit, i.e., to receive and irreversibly record one bit, is given by Landauer's Principle as $\ln 2 \ k_B T$, with $k_B$ Boltzmann's constant and $T$ temperature [58, 59, 60]. The (minimum) time to irreversibly obtain one bit is then $\hbar/\ln 2 \ k_B T$, roughly 30 fs at 310 K. For comparison, the thermal dissipation time (in 3d space) due to time-energy uncertainty is $\pi\hbar/2\ln 2 \ k_B T$ [61], roughly 50 fs at 310 K. These values define a minimal timescale for biologically-relevant, irreversible information processing, roughly the timescale of molecular-bond vibrational modes [62] and an order of magnitude shorter than photon-capture timescales [63].

Viewing all physical interaction as information exchange – and the bit as the fundamental "quantum" of information – has the immediate consequence that interaction discretizes the state spaces of all observable degrees of freedom. Given a set of mutually-commuting, binary-valued observables, i.e., quantum operators implementing yes/no questions as described below, a discrete state space for any observed system is constructed by assigning a basis vector to each possible outcome (yes or no, +1 or -1) for each observable. These are finite-dimensional Hilbert spaces. While Hilbert spaces with either finite or infinite

6

dimension are introduced *ad hoc* in traditional quantum mechanics [24], finite-dimensional Hilbert spaces emerge naturally from the process of observation in the information theory. As Fuchs puts it, infinite dimensional Hilbert spaces are, from an information perspective, merely a "useful artifice" permitting computation with differential equations [64]. Hence in what follows, all Hilbert space dimensions will be finite dimensional.

## 2.2 Unitarity

The fundamental axiom of quantum theory is unitarity, again introduced *ad hoc* in traditional treatments. If $U$ is an isolated or *closed* system, its time propagator is a unitary operator:

$$\mathcal{P}_U = e^{-(\imath/\hbar)H_U t}, \tag{1}$$

where $H_U$ is the "internal" Hamiltonian (i.e. energy) operator satisfying the Schrödinger equation:

$$\imath\hbar(\partial/\partial t)|U(t)\rangle = H_U|U(t)\rangle, \tag{2}$$

where $|U(t)\rangle$ is the time-dependent state of $U$. When $H_U$ and hence $\mathcal{P}_U$ are time-invariant, solutions have the form:

$$|U(t)\rangle = e^{-\imath\varphi t}|U(0)\rangle, \tag{3}$$

where $|U(0)\rangle$ is an initial ($t = 0$) state. This Eq. (3) describes a phase rotation by $\varphi$ per unit time $t$ in the Hilbert space $\mathcal{H}_U$; the initial state $|U(0)\rangle$ is preserved "up to" this phase rotation. The dimension $d = \dim(\mathcal{H}_U)$ is the number of basis vectors of $\mathcal{H}_U$ and is, by definition, the quantity (in bits) of *observable* information encoded by the state $|U(0)\rangle$. The number $d$ is clearly invariant under the phase rotation given by Eq. (3); hence the phase rotation is not an observable. Unitary evolution as defined by Eq. (1) is, therefore, simply evolution over time that conserves observable information. Hence, the fundamental axiom of quantum theory is not *ad hoc* at all: it is the Principle that observable information, like energy, is neither created nor destroyed by physical processes.

The appearance of the "imaginary" unit $\imath$ in Eq. (1) - (3) and the use of Hilbert spaces over the complex field $\mathbb{C}$ are standard in quantum theory but are often considered a mere convenience; compelling arguments for their necessity have only been developed recently [65, 66]. They can, however, be given a straightforward interpretation: they emphasize that the phase rotation implemented by $\mathcal{P}_U$ is not an observable dynamics and that the "external" time $t$ in these equations is not an observable, clock-referenced time.[2] Indeed, no classical

---

[2]Imagine, for an example, listening to a constant tone that has no beginning or end. Writing time as $\imath t$ with $t$ real allows Minkowski spacetime to be given a Galilean $(++++)$ metric. See [67] for a discussion of $\imath$ as a shorthand for converting classical to quantum observables, and [68] for a more intuitive view of "rotation by $i$" as a formal operation.

information – no observational outcome – has yet been obtained in the setting defined by Eq. (1) - (3). Characterizing observation as a process generating classical outcomes as a result requires an additional assumption of separability as outlined below.

## 2.3 Separability and holographic encoding

Let $U$ be an isolated system as above, and let $U = AB$ be a bipartite decomposition of $U$ into proper components, i.e. $A \neq U$ and $B \neq U$. The components $A$ and $B$ are complements within the "universe" $U$; $B$ comprises all of $U$ that is not $A$ and vice-versa. The system $B$ includes, in particular, any "systems" with which $A$ interacts as well as the "environment" that surrounds them, as discussed in detail in §3 below. The joint Hilbert space is then $\mathcal{H}_U = \mathcal{H}_A \otimes \mathcal{H}_B$. The components $A$ and $B$ are in this case, by definition, *open* systems; $A$ is open to energy and information exchange with $B$ and vice versa. At a fixed time $t$, any such bipartite decomposition can be characterized by an entanglement entropy:

$$\mathcal{S}(|AB\rangle) = -\sum_i |\alpha_i|^2 \log_2(|\alpha_i|^2), \tag{4}$$

where the coefficients $\alpha_i$ are the Schmidt coefficients given by:

$$|AB\rangle = \sum_i^m \alpha_i |u_i\rangle_A |v_i\rangle_B, \tag{5}$$

where the label $B$ is assigned so that $m = \dim(\mathcal{H}_B) \leq \dim(\mathcal{H}_A)$ and the $|u_i\rangle_A$ and $|v_i\rangle_B$ are orthonormal states of $A$ and $B$ respectively. The entanglement entropy $\mathcal{S}(|AB\rangle)$ is a mutual information measure that detects quantum correlation or "coherence" between $A$ and $B$. If $\mathcal{S}(|AB\rangle) = 0$, the joint state factors as $|AB\rangle = |A\rangle|B\rangle$ and hence is *separable*; otherwise the joint state is *entangled*. Separable states are also called *decoherent*; entangled states are *coherent*. Heuristically, a joint state $|AB\rangle$ is separable if the individual states $|A\rangle$ and $|B\rangle$ can be determined by independent measurements. If $|AB\rangle$ is entangled, interactions with $A$ and $B$, including measurements, are no longer independent; this lack of independence can be detected [56] and is the empirical basis for demonstrating entanglement [69] as discussed in §5.4 below.

Extending Eq. (4) to all times and all bipartite decompositions of $U$ gives a representation of the time-dependent entanglement structure of $U$. As $t \to \infty$ the unitary dynamics of Eq. (1) will drive $U$ toward maximal entanglement, i.e. $\mathcal{S}(|AB\rangle) \to \dim(\mathcal{H}_B)$ for every bipartite decomposition $U = AB$; where, here again, $B$ is taken to be the smaller of the two systems. At maximum entanglement, the joint-state evolution can be considered a simple rotation as in Eq. (3).

Given a decomposition $U = AB$, the Hamiltonian can be decomposed as $H_U = H_A + H_B + H_{AB}$, where $H_{AB}$ represents the $A$-$B$ interaction. Provided $H_{AB}$ is weak compared to the internal interactions $H_A$ and $H_B$ and the time period of interest is short compared to

timescale in which $\mathcal{P}_U$ drives the joint system to maximal entanglement, $A$ and $B$ can be considered at least approximately separable, i.e. $\mathcal{S}(|AB\rangle) \approx 0$ so $|AB\rangle \approx |A\rangle|B\rangle$. In this case, $|A\rangle$ and $|B\rangle$ can be considered individually well-defined, and bases can be chosen for $A$ and $B$ so that:

$$H_{AB} = \beta^k k_B T^k \sum_i^N \alpha_i^k M_i^k, \tag{6}$$

where $k = A$ or $B$, the $M_i^k$ are $N$ Hermitian operators with eigenvalues in $\{-1, 1\}$, the $\alpha_i^k \in [0, 1]$ are such that $\sum_i^N \alpha_i^k = 1$, and $\beta^k \geq \ln 2$ is an inverse measure of $k$'s thermodynamic efficiency that depends on the internal dynamics $H_k$ [46, 47, 70, 71]. For fixed $k$, the operators $M_i^k$ clearly must commute, i.e. $[M_i^k, M_j^k] = M_i^k M_j^k - M_j^k M_i^k = 0$ for all $i, j$; hence $H_{AB}$ is swap-symmetric under the permutation group $S_N$ for each $k$. The thermodynamic factor $\beta^k k_B T^k$ in Eq. (6) assures compliance with Landauer's Principle, i.e., assures that the per-bit free-energy cost of classical bit erasure is paid on each cycle (see [46, 47] for discussion). As $U = AB$ is by assumption isolated, conservation of energy requires $\beta^A T^A = \beta^B T^B$. As discussed in [13], Eq. (6) can be written in ordinary narrative form as:

$$\text{Physical Interaction} = (\text{Thermodynamics}) \cdot (\text{Yes/No questions}).$$

This formulation emphasizes what quantum theory is about: the process of obtaining information. Obtaining information from $B$ requires, in particular, that $A$ acts on $B$ by asking questions. As Wheeler [72] puts it, "No question? No Answer!" All inference in this framework is active inference; Eq. (6) does not allow "passive perception" to be a physical process.

In contrast to classical theories of information transfer, in which physical tokens encoding bits are transmitted between observers at different locations, Eq. (6) involves no assumptions about spacetime, objects, or motions. It is strictly topological: given separability, it identifies a boundary $\mathscr{B}$ between $A$ and $B$ at which $H_{AB}$ is defined. This boundary is the "channel" via which $A$ and $B$ exchange strings of bits, ordered by the order of the operators $M_i^k$ in Eq. (6). Each of these bit strings encodes one eigenvalue of $H_{AB}$; as $H_{AB}$ has units of energy, these eigenvalues are measures of the energy $\beta^k k_B T^k$ exchanged between $A$ and $B$ during each cycle of interaction. Any time variation of $H_{AB}$ is, therefore, time variation of the energy exchanged and can be written:

$$H_{AB}(t) = \beta^k(t) k_B T^k \sum_i^N \alpha_i^k(t) M_i^k, \tag{7}$$

where $\alpha_i^k(t)$ and $\beta^k(t)$ are subject to the conditions on $\alpha_i^k$ and $\beta^k$ given above. As noted for Eq. (1) - (3) above, the time $t$ in Eq. (7) is not an observable, clock-referenced time. We will for simplicity consider $H_{AB}$ to be $t$-invariant; this is effectively an adiabatic assumption. An observer-specific, measurable time will be introduced in §3.1 below.

Boundaries such as $\mathscr{B}$ that function as information channels are, when embedded in space-time and constrained by general covariance,[3] *holographic screens* that limit communication between the regions they separate to the bits that they encode [73, 74, 75, 76]. Interactions between separable systems, i.e. interactions of the form given by Eq. (6) or (7) can, without loss of generality, be regarded as defined at such holographic screens [47, 71, 77]. The operators $M_i^k$ can, in this case, be regarded as "preparation" and "measurement" operators that alternately write and read bit values encoded on $\mathscr{B}$. The $S_N$ swap symmetry of the $M_i^k$ for each $k$ means that the bits can be prepared, and then measured, in any order and hence independently, provided preparation precedes measurement for each bit (such swaps change the "zero point" of the energy scale and are undetectable). The screen $\mathscr{B}$ is, as is any holographic screen, an ancillary construct, not "part of" either $A$ or $B$. It can be physically realized as an ancillary array of noninteracting qubits (quantum bits) with which $A$ and $B$ interact as illustrated in Fig. 1.

---

[3]Effectively, this is a requirement for consistency with both Special and General Relativity. Note that we employ the Holographic Principle (HP) in its original sense as defined by 't Hooft [73] and Susskind [74] and elaborated by Bousso [75]: the HP places an upper bound on the classical, thermodynamic entropy of any system as measured by an observer outside that system. The term "holographic principle" is sometimes used to denote the holographic duality between a gravitational theory in N dimensional Anti de Sitter (AdS) space and a conformal (generally quantum) field theory (CFT) on its N-1 dimensional boundary. This AdS/CFT duality is a special case of the HP.
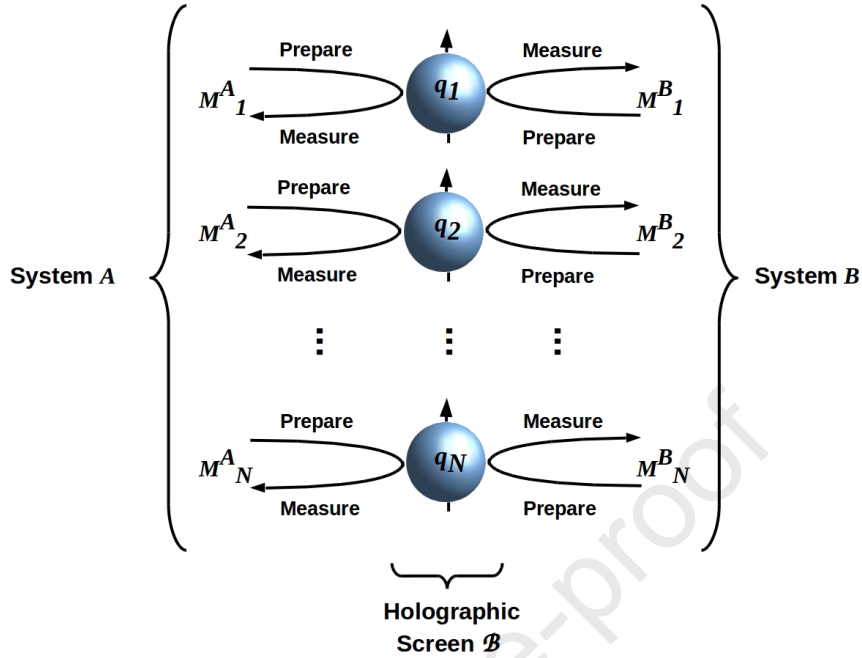
Figure 1: A holographic screen $\mathscr{B}$ separating systems $A$ and $B$ with an interaction $H_{AB}$ given by Eq. (6) can be realized by an ancillary array of noninteracting qubits that are alternately prepared by $A$ ($B$) and then measured by $B$ ($A$). Qubits are depicted as Bloch spheres [36]. There is no requirement that $A$ and $B$ share preparation and measurement bases, i.e. QRFs. Adapted from [47] Fig. 1, CC-BY license.

The holographic screen $\mathscr{B}$ has an obvious interpretation in the language of the FEP and active inference: it implements the MB separating $A$ from $B$ [78]. "Active" and "sensory" states of the MB correspond to preparation and measurement implemented by the $M_i^k$; as the interaction $H_{AB}$ is perfectly symmetrical, $A$'s actions are $B$'s sensations and vice-versa. As bit strings on $\mathscr{B}$ encode energy eigenvalues, what either $A$ or $B$ "senses" is energy. Indeed any physical interaction, when framed in terms of scattering, can be viewed as an exchange of energy across a holographic screen functioning as an MB [47], all path degrees of freedom are effectively encapsulated by the MB. The assumption of separability plays, in this setting, the role played by the assumption of measurability in [10]: it guarantees that systems $A$ and $B$ have well-defined individual states. These states are at *informational* equilibrium; $A$ and $B$ exchange bits one-for-one. They are not, however, at thermal equilibrium, $T^A = T^B$ unless their thermodynamic efficiencies $\beta^A = \beta^B$. Hence, when sampled and averaged over macroscopic times, the pure states $|A\rangle$ and $|B\rangle$ become (mixed) NESS densities $\rho_A$ and $\rho_B$.

11

## 2.4  Reference frames

The isolation of $U = AB$ assures that the $A - B$ channel is free of classical, environmentally-induced noise; as $U$ is isolated, it has, by definition, no "environment" from which noise could flow. The component systems $A$ and $B$ are, however, open to each other as noted above. The preparation and measurement operations $M_i^A$ and $M_i^B$ that $A$ and $B$ use to communicate are defined with respect to and hence depend on the bases $|u_i\rangle$ and $|v_i\rangle$ of the Hilbert spaces $\mathcal{H}_A$ and $\mathcal{H}_B$ respectively. In the qubit realization shown in Fig. 1, the $M_i^A$ and $M_i^B$ are $z$-spin operators and so depend on the "choices" of $z$ axis $z_A$ and $z_B$. Provided $A$ and $B$ are separable, and assuming that there are no "superdeterminist" *a priori* correlations, $z_A$ and $z_B$ are uncorrelated; this is the "free choice" assumption. Free choice is often claimed to be essential to science as a practice [79, 80]; if it characterizes any bounded system, consistency with quantum theory and special relativity together requires that it must characterize all such systems [81][4]. Free choice of a basis introduces quantum "noise" that is indistinguishable, by either $A$ or $B$, from classical noise. Suppose $B$ encodes $|\uparrow\rangle$ on qubit $q$, using $z_B$ to define the "up" direction. If $z_A = z_B$, $A$'s measured state $|q\rangle_A = |\uparrow\rangle$, i.e. $A$'s probability $P_A(|\uparrow\rangle) = 1$. If, however, $z_A$ is chosen perpendicular to $z_B$, $|q\rangle_A = (|\uparrow\rangle - |\downarrow\rangle)/\sqrt{2}$ and $P_A(|\uparrow\rangle) = P_A(|\downarrow\rangle) = 0.5$.

The $z$ axis in Fig. 1 is a *reference frame*; free choice of the $z$ axis generalizes to free choice of the reference frame for encoding each qubit on $\mathscr{B}$. While in classical physics, reference frames are typically thought of as fully-specifiable abstractions, in quantum theory reference frames must be considered physical systems – QRFs – that encode unmeasurable quantum phase information. A QRF cannot, therefore, be fully specified by any finite bit string; it is "nonfungible" in the terminology of [49]. In the setting of Fig. 1, the $z$ axis deployed by a $A$ ($B$) determines how $A$ ($B$) prepares and then measures the states of the qubits composing $\mathscr{B}$. We can, in particular, consider $A$ to be isolated from $B$ during the time interval "between" preparation and measurement steps, an interval shorter than the natural timescale of the interaction $H_{AB}$. During such a short interval, we abuse the notation only slightly by writing:

$$\mathcal{P}_k : |\mathscr{B}\rangle_{meas} \mapsto |\mathscr{B}\rangle_{prep} \tag{8}$$

where $k = A$ or $B$, i.e. by thinking of the internal propagators $\mathcal{P}_A$ and $\mathcal{P}_B$, and hence the dynamics $H_A$ and $H_B$, as computing the next preparation of the qubits on $\mathscr{B}$ from their most recent measured state. The idea of "computation" – or more properly, of the physical

---

[4]The free choice assumption relates to the philosophical concept of free will via the Conway-Kochen "free will" theorem [81], which states, under generic assumptions, that if any physical system has free choice, all physical systems do. "Free choice" here means freedom from *local* determination, i.e. from determination by events in the system's past light cone. In a universe lacking singularities (e.g. classical "hard spheres" that instantaneously alter momenta), local determinism is inconsistent with global or superdeterminism. Tipler [82] has shown that removing the singularities from globally-determinist classical physics yields (Bohmian) quantum theory with no further assumptions. Freedom from local determinism for all systems entails that no local system can exercise local determinism [83]; all local actions have, in other words, unpredictable side effects.

system $A$ ($B$) *implementing* a computation – is the standard notion, reviewed in [84]: in brief, we can say a system $A$ implements a computation of a function $\mathbf{F}$ if an "interpretation" map $\mathcal{I}_{\mathbf{F}}$ exists such that the diagram:

$$
\begin{array}{ccc}
\langle\text{bits}\rangle_i & \xrightarrow{\ \mathbf{F}\ } & \langle\text{bits}\rangle_{i+1} \\
\mathcal{I}_{\mathbf{F}}\Big\uparrow & & \Big\uparrow\mathcal{I}_{\mathbf{F}} \\
|\mathscr{B}\rangle_i & \xrightarrow[\ \mathcal{P}_A\ ]{} & |\mathscr{B}\rangle_{i+1}
\end{array}
\tag{9}
$$

commutes, i.e. $\mathbf{F}\,\mathcal{I}_{\mathbf{F}} = \mathcal{I}_{\mathbf{F}}\mathcal{P}_A$ for every "step" $i \to i+1$, where here $\langle\text{bits}\rangle$ is a finite bit string and $i \to i+1$ generalizes $meas \to\ prep$ in Eq. (8). A QRF, e.g. a $z$ axis as in Fig. 1, implemented by $\mathcal{P}_A$ effectively "chooses" the computational basis that renders the string $\langle\text{bits}\rangle$ well-defined. This is the same choice of basis that renders the $M_i^A$ well-defined in Eq. (6). Hence, we can identify the interpretation map $\mathcal{I}_{\mathbf{F}}$ with the QRF that renders the function $\mathbf{F}$ well-defined; to simplify the notation, we will use boldface $\mathbf{F}$ to label the QRF itself. The choice of QRF $\mathbf{F}$ and hence of computational basis is *functional* or *semantic*, not physical; both $H_A$ and $H_{AB}$ are invariant under changes of basis for $\mathcal{H}_A$. It is worth noting explicitly that the choice of $\mathbf{F}$ is not encoded on $\mathscr{B}$ and is not observationally accessible to $B$. Indeed, the general question of whether an arbitrary system $A$ implements a QRF $\mathbf{F}$ is Turing undecidable by Rice's theorem [85]; see [47] for discussion.

As illustrated by the $z$ axes in Fig. 1, the role of any QRF in a measurement setting is to assure that measured values of some degree of freedom can be compared with each other and hence given an operational meaning [86]. Meter sticks, clocks, gyroscopes, and the Earth with its gravitational and magnetic fields serve this function, and are commonplace laboratory QRFs. As described in detail in §3.3 below, recognizing an external object, such as a meter stick, as a QRF and using it as such requires that the observer in question already implements a QRF for the relevant degree of freedom: an observer with no "internal" ability to measure and compare lengths, for example, would find a meter stick useless. All QRFs can, therefore, be considered internal functional components of, or computations implemented by, larger systems that allow measurements of, and assign operational semantics to, one or more degrees of freedom external to the implementing system [70]. Shared operational semantics across observers requires shared QRFs. The nonfungibility of QRFs renders the question of QRF sharing in general Turing undecidable, again by Rice's theorem [47], a result consistent with the general undecidability of language sharing [87].

## 2.5  Symmetry breaking, decoherence, and agency

We will be interested in what follows in open quantum systems, considered to be observers, that deploy one or more distinct QRFs to measure particular subsets of the bits encoded on their boundaries/MBs, and that record the values of these bits to a memory that persists for at least one measurement cycle and hence allows comparisons of the values obtained in at least two sequential measurements. Such systems effectively decompose the holographic

screen $\mathscr{B}$ into three disjoint sectors that we will label $E$ the *observed environment*, $F$ the *unobserved environment*, and $Y$ the *memory* sector, respectively. Equivalently, they effectively decompose the set $M_i^k$ of operators into three disjoint subsets (dropping the redundant index $k$) $M_i^E$, $M_j^F$, and $M_l^Y$, respectively, with:

$$X =_{def} \text{dom}(\{M_i^X\}) \tag{10}$$

for each sector $X$. Assuming free choice of basis as above, this decomposition into sectors can be regarded as freely chosen. Note that as $\mathscr{B}$ is the only locus of classical information in the current formalism, any persistent classical memory *must* be a sector on $\mathscr{B}$ as discussed in [13].

The assignment of operators to specific sectors breaks the $S_N$ swap symmetry of $\mathscr{B}$ [47]. As with choice of QRF, this is a functional or semantic symmetry breaking, not a physical symmetry breaking; the assignment of bits on $\mathscr{B}$ to distinct sectors by $H_A$ has no effect on the definition of $H_{AB}$ given by Eq. (6). Holding $H_{AB}$ fixed, we can vary the internal interaction $H_A$, and hence the implementation of QRFs as computations, so as to either satisfy $S_N$ symmetry or break it. Note that varying $H_A$ in this way is equivalent to varying $H_U - H_B$; such variation is undetectable at the boundary $\mathscr{B}$. This insensitivity of $\mathscr{B}$ to variation in the internal or joint dynamics of $A$ and $B$ is the core physical meaning of the holographic principle [73, 74, 75, 76].

Breaking the $S_N$ swap symmetry on $\mathscr{B}$ renders the sector states $|E\rangle$, $|F\rangle$, and $|Y\rangle$, each of which corresponds to an encoded bit string, separable and hence mutually decoherent [47]. As emphasized in [88, 89, 90, 91, 92, 93] among others, decoherence is always observer-relative, even when the "observer" is an ambient environment. Hence decoherence due to swap-symmetry breaking is decoherence *relative to* $A$; the change-of-basis invariance of $H_{AB}$ renders $A$'s sector boundaries undetectable by $B$. Given free choice of QRFs, moreover, $B$'s sector boundaries, if any, may be different from $A$'s. Mismatched sector boundaries between $A$ and $B$ generate apparent "hidden variables" and hence variational free energy [4, 5, 6, 10] as discussed in §4.2 below. For the present purposes, what is important is that deploying a QRF sensitive to only some degrees of freedom of $\mathscr{B}$ (i.e. sensitive to only some bit values encoded on $\mathscr{B}$) breaks the swap symmetry on $\mathscr{B}$ and creates a decoherent sector as defined by Eq. (10).

Breaking the swap symmetry on $\mathscr{B}$ allows different sectors to have different thermodynamic efficiences, i.e. breaking swap symmetry allows breaking thermodynamic symmetry. This can be made evident by rewriting Eq. (6) for $A$ in terms of sectors $X$ as:

$$H_{AB} = \sum_X \beta^X k_B T^A \sum_{i \in i_X} \alpha_i^A M_i^A, \tag{11}$$

where $i_X$ is the set of indices of operators $M_i^A$ assigned to sector $X$. Note that this symmetry breaking does not change the total energy exchanged (i.e. the eigenvalue of $H_{AB}$), it only allocates the energy differently to different sectors. This allows sectors to perform different thermodynamic roles; in particular, it allows the unobserved environment sector

$F$ to serve as a source of free energy to – and a sink of waste heat from – the observed environment sector $E$ and the memory sector $Y$ [46, 47]. The bits encoded on $F$ are, therefore, noninformative to $A$ and are traced (effectively, marginalised and averaged) over when defining information-bearing states of $A$. Mathematically, this trace operation can be viewed as implementing decoherence as discussed in [25, 26, 27, 28, 29], i.e., as implementing the sector boundary of $F$. Assigning this thermodynamic function to $F$ enables thermodynamically-expensive classical information processing of bits encoded on $E$ and $Y$, including maintaining the stability of bit values written to $Y$ as discussed in §3.1 below. It therefore allows $E$ and $Y$ to have distinct semantics.

Free choice of a decomposition of $\mathscr{B}$ into sectors with different QRF-induced semantics is indicative of agency. Hence for the purposes of this paper, we can define:

**Definition 1.** *A (nontrivial) agent is a system $A$ with an internal dynamics $H_A$ that breaks the $S_N$ swap symmetry of its boundary/MB $\mathscr{B}$.*

We can consider a system with a swap-symmetric boundary as shown in Fig. 1 a *trivial* agent. Trivial agents correspond to "inert" systems with functionally-insignificant internal states, and hence no internal information-processing capacity, as discussed in [10]; to employ an example from this paper, a rock is a trivial agent, because it does not allocate inputs from different parts of its surface to different information-processing pathways. A bacterium, on the other hand, does do this; hence bacterial chemotaxis is both possible and effective. All nontrivial agents engage in active inference as shown below; hence they behave so as to increase their subjective (i.e. agent-relative) probability of their own future existence. We consider nontrivial agents to be "cognitive" in at least a basal sense; see [13] for discussion.[5]

Before proceeding to further characterize these sectors or to consider the imposition of additional structure on $E$ by further QRFs, we briefly review below the specification of QRFs using the generic formalism of Channel Theory [52]. This formalism will enable characterization of the asymptotic behavior of the FEP and of context-switching as a mechanism for minimizing FEP as discussed below in §4.3 below.

## 2.6 Channel theory of QRFs

Channel Theory [52] is an application of the category theory of Chu spaces, spaces of semantic relations exemplified by object – attribute tables [95, 96, 97]. Indeed Channel

---

[5]The words "agency" and "agent" are used here in a deflationary sense; namely, that to infer, measure or exchange renders a system and agent of inference, measurement or exchange, in a non-propositional sense. Different (natural) kinds or levels of agency - such as that entailed by planning and minimal selfhood - are left for different treatment. For example, we have deliberately suppressed finer details to foreground the basic ideas. One finer detail found in the FEP literature, is the distinction between active and sensory states.that jointly constitute blanket states. It could be argued that Markov blankets which do not possess active states stipulate things that "exist" but have no "agency". In this regard, we expect much to be related to the phenomenology of bio-cognitive sensorimotor agency as surveyed in e.g. [94] (and see references therein).

Theory can be regarded as defining a category **Chan** that is isomorphic to the category **Chu(Set**, $K$) (for short, **Chu**) of Chu spaces. While conceptually simple, Channel Theory is surprisingly rich, providing both a natural representation of conditional probabilities and formal criteria for operator commutativity and quantum contextuality [98, 99] (for a logico-philosophical perspective on the properties of semantic coherence in **Chan**, see e.g. [100]). Furthermore, it provides an implementation-independent formal language for writing functional specifications of QRFs as computations.

The central idea of Channel Theory is that of a "classifier" that relates tokens in some language to types in that language. We can define a classifier as an object in **Chan** as follows:

**Definition 2.** *A* classifier $\mathcal{A}$ *is a triple* $\langle Tok(\mathcal{A}), Typ(\mathcal{A}), \models_{\mathcal{A}} \rangle$ *where* $Tok(\mathcal{A})$ *is a set of "tokens",* $Typ(\mathcal{A})$ *is a set of "types", and* $\models_{\mathcal{A}}$ *is a "classification" relating tokens to types.*

The classification $\models_{\mathcal{A}}$ can, in general, be valued in any set $K$ without assumed structure (as is the case for Chu spaces); for simplicity, we will consider only binary classifications. Morphisms in **Chan**, called "infomorphisms" between these objects are then given by the following:

**Definition 3.** *Given two classifiers* $\mathcal{A} = \langle Tok(\mathcal{A}), Typ(\mathcal{A}), \models_{\mathcal{A}} \rangle$ *and* $\mathcal{B} = \langle Tok(\mathcal{B}), Typ(\mathcal{B}), \models_{\mathcal{B}} \rangle$, *an* infomorphism $f : \mathcal{A} \to \mathcal{B}$ *is a pair of maps* $\overrightarrow{f} : Tok(\mathcal{B}) \to Tok(\mathcal{A})$ *and* $\overleftarrow{f} : Typ(\mathcal{A}) \to Typ(\mathcal{B})$ *such that* $\forall b \in Tok(\mathcal{B})$ *and* $\forall a \in Typ(\mathcal{A})$, $\overrightarrow{f}(b) \models_{\mathcal{A}} a$ *if and only if* $b \models_{\mathcal{B}} \overleftarrow{f}(a)$.

This last definition can be represented schematically as the requirement that the following diagram commutes:

$$
\begin{array}{ccc}
\mathrm{Typ}(\mathcal{A}) & \xrightarrow{\overleftarrow{f}} & \mathrm{Typ}(\mathcal{B}) \\
{\scriptstyle\models_{\mathcal{A}}}\Big| & & \Big|{\scriptstyle\models_{\mathcal{B}}} \\
\mathrm{Tok}(\mathcal{A}) & \xleftarrow{\overrightarrow{f}} & \mathrm{Tok}(\mathcal{B})
\end{array}
\tag{12}
$$

An infomorphism $f : \mathcal{A} \to \mathcal{B}$ is, effectively, a map relating the semantic constraints imposed by the classification $\models_{\mathcal{A}}$ to those imposed by $\models_{\mathcal{B}}$.

We are, in practice, interested in collections of infomorphisms than construct complex semantic constraints out of simple ones; these will allow us to specify QRFs as hierarchies of semantic constraints. Given a finite collection $\mathcal{A}_i$ of classifiers, we can represent this construction process as a finite, commuting *cocone diagram* (CCD) depicting a flow of infomorphisms sending inputs to a *core* $\mathbf{C}'$ that is the category-theoretic colimit of the underlying classifiers, i.e., is the apex of the maximally general diagram of this form over the $\mathcal{A}_i$, if a unique such maximum exists:

$$
\begin{array}{c}
\mathbf{C}' \\
{\scriptstyle f_1}\nearrow \quad {\scriptstyle f_2}\big\uparrow \quad \nwarrow{\scriptstyle f_k} \\
\mathcal{A}_1 \xrightarrow{\;g_{12}\;} \mathcal{A}_2 \xrightarrow{\;g_{23}\;} \ldots \mathcal{A}_k
\end{array}
\tag{13}
$$

16

The cocone core $\mathbf{C}'$ is itself a classifier that encodes, via the incoming infomorphisms $f_i$, the conjunction of the semantic constraints imposed by the $\mathcal{A}_i$.

There is a dual construction to this CCD, namely a commuting finite *cone diagram* (CD) of infomorphisms on the same classifiers, where all arrows are reversed. In this case the core of the (dual) channel is the category-theoretic limit of all possible downward-going structure-preserving maps to the classifiers $\mathcal{A}_i$. Hence, we can define the central idea of a finite, commuting *cone-cocone diagram* (CCCD) as consisting of both a cone and a cocone on a single finite set of classifiers $\mathcal{A}_i$ linked by infomorphisms as depicted below:

$$
\begin{array}{c}
\end{array}
\tag{14}
$$

If the cores $\mathbf{C}' = \mathbf{D}'$, we can also represent the CCCD as:

$$
\tag{15}
$$

This diagram is naturally interpreted as reconstructing the semantics of the $\mathcal{A}_i$ via the "combined" representation $\mathbf{C}'$. Generalizing Diagram (15) by letting $\mathbf{C}'$ be the limit of a smaller set of classifiers $\mathcal{A}'_1, \ldots \mathcal{A}'_j$, $j < k$, we can write:

$$
\tag{16}
$$

Diagram (16) provides a natural representation of coarse-graining the semantics of $\mathcal{A}_i$ via $\mathbf{C}'$ into a compressed representation $\mathcal{A}'_i$. We will employ this generalization in §3.1 below to specify the writing of coarse-grained records of observational outcomes to the memory sector $Y$.

Diagrams such as (13) – (16) can be generalized into hierarchical networks by adding intermediate layers of classifiers and appropriate maps; when this is done, they clearly

17

resemble artificial neural networks (ANNs), and in the "bowtie" form of Diagrams (15) and (16), variational autoencoders (VAEs) [98]. The core $\mathbf{C}'$ in (15) and (16) can be viewed as both an "answer" computed by the $f_i$ from inputs to the $\mathcal{A}_i$ and, dually, as an "instruction" propagated by the $h_i$ (or in Diagram (16), the $h_i'$), to drive outputs from the $\mathcal{A}_i$ (or in Diagram (16), the $\mathcal{A}_i'$). Such dual input/output behavior is exactly the behavior of a QRF. We can, therefore, represent any QRF as a CCCD "attached" to a subset of measurement operators $M_k^A, \ldots M_n^A$ by maps that identify the binary eigenvalues of the $M_i^A$ with binary inputs to the $\mathcal{A}_i$ as illustrated in Fig. 2. This category-theoretic representation is particularly useful in the present context, as it is completely independent of the physical implementation of the relevant QRF.
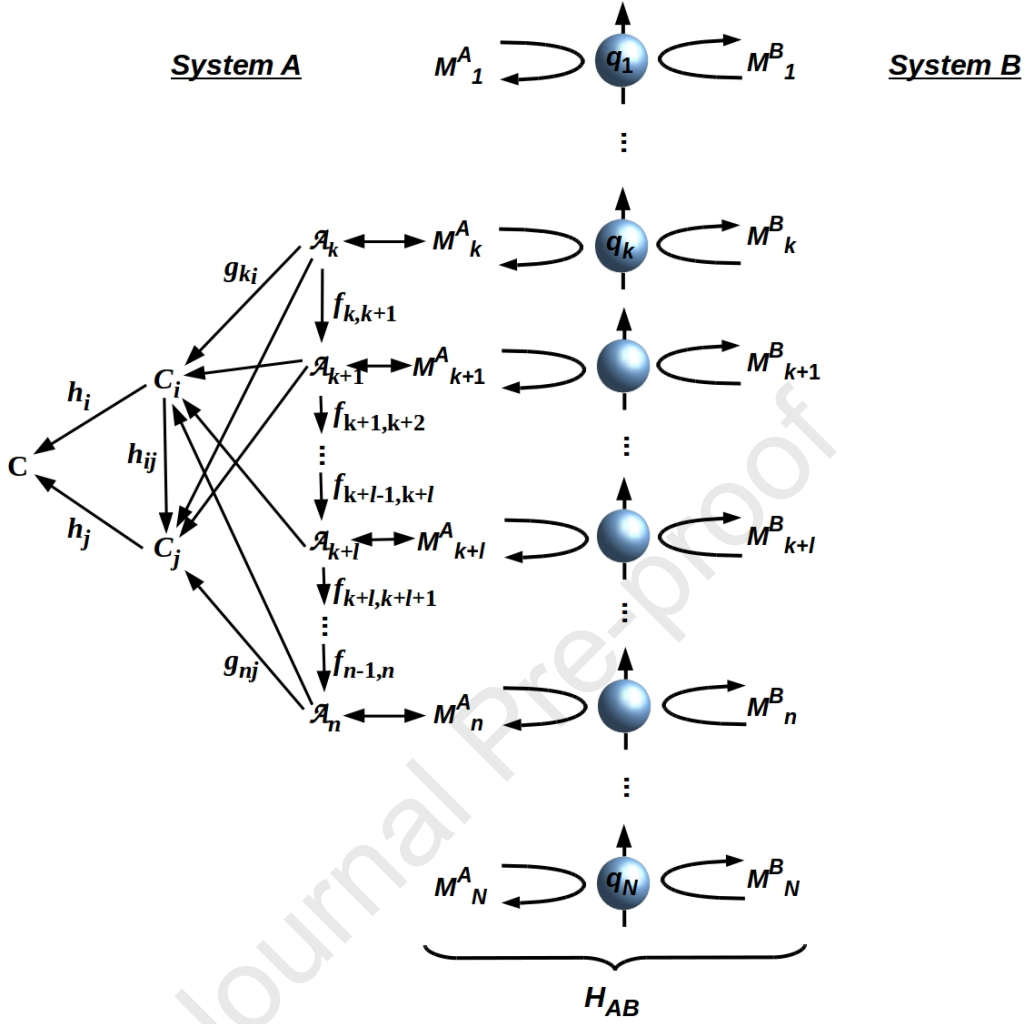
Figure 2: Attaching a CCCD to a subset of measurement operators $M_k^A, \ldots M_n^A$ by identifying the binary eigenvalues of the $M_i^A$ with binary inputs to the $\mathcal{A}_i$. Only the CCD direction arrows are shown for simplicity; adding equivalent but reversed arrows completes the CCCD. The CCCD specifies a function computed by the internal dynamics $\mathcal{P}_A$, i.e. a QRF deployed by $A$. Adapted from [47] Fig. 3; CC-BY license.

Thus far we have considered binary CCCDs, corresponding to Boolean ANNs or VAEs or to QRFs imposing Boolean constraints. Mapping a CCCD to an arbitrary ANN or VAE that computes some function of interest $\mathbf{F}$ just requires assigning probabilities, i.e. "weights" to the infomorphisms in a way that respects the Kolmogorov axioms [98, 99]. We can then treat the computed function $\mathbf{F}$ from network inputs to network outputs as an arbitrary probabilistic QRF, or as suggested by Diagram (16), a pair of coupled QRFs that process some "sensory" signal received by the $\mathcal{A}_i$ and write the result to a lower-dimensional and

hence coarse-grained "memory" via the $\mathcal{A}'_i$. A QRF so defined is naturally interpreted as performing hierarchical Bayesian inference [98, 99]; see [101, 102] for applications to human cognition.

To assign probabilities to infomorphisms, it is convenient to add the structure of a "local logic" $\mathcal{L}(\mathcal{A})$ relating subsets of tokens and types to each classifier $\mathcal{A}$ [52]. To do this, we define an "implication" relation between subsets of types:

**Definition 4.** *Two subsets* $M, N \subseteq Typ(\mathcal{A})$ *are related by a* sequent $M \models_{\mathcal{A}} N$ *if* $\forall x \in Tok(\mathcal{A}), x \models_{\mathcal{A}} M \Rightarrow x \models_{\mathcal{A}} N$.

Via the sequent relation, the local logic $\mathcal{L}(\mathcal{A})$ effectively arranges the types of $\mathcal{A}$ into a hierarchy; any classifier can be extended to a local logic in this way by adding types as needed. Considering all classifiers to be extended to local logics, Diagrams (13) – (16) can be considered diagrams of local logics by requiring each of the infomorphisms to be a "logic infomorphism" that preserves the sequent structure. In general, given an information flow channel:

$$\longrightarrow \mathcal{A}_{\alpha-1} \longrightarrow \mathcal{A}_{\alpha} \longrightarrow \mathcal{A}_{\alpha+1} \longrightarrow \cdots \tag{17}$$

the semantic content can be extended by postulating local logics $\mathcal{L}_{\alpha} = \mathcal{L}(\mathcal{A}_{\alpha})$ generated by the corresponding classifiers $\mathcal{A}_{\alpha}$ (assumed, in principle, to be in relationship to a (regular) theory associated to the individual $\mathcal{A}_{\alpha}$, as specified in [52, Ch 9] (cf. [103]) and reviewed in [99, Appendix A]), so to obtain a flow of logic infomorphisms:

$$\cdots \longrightarrow \mathcal{L}_{\alpha-1} \longrightarrow \mathcal{L}_{\alpha} \longrightarrow \mathcal{L}_{\alpha+1} \longrightarrow \cdots \tag{18}$$

which we can take to comprise comprise a CCD as in Eq. (13). Logic infomorphisms are, effectively, embeddings of type hierarchies; Diagrams (15) and (16) can be viewed as embeddings into a "top-level" type hierarchy $\mathbf{C}'$ that assigns an overall semantics to its inputs, followed by encodings of this top-level hierarchy into some componential representation.

The local logic $\mathcal{L}(\mathcal{A})$ defined above is Boolean. To extend the type hierarchy defined by $\mathcal{L}(\mathcal{A})$ to a hierarchical Bayesian inference, we extend $\mathcal{L}(\mathcal{A})$ to a probabilistic logic by relaxing the sequent relation to require only that if $x \models_{\mathcal{A}} M$, there is some probability $P(N|M)$ that $x \models_{\mathcal{A}} N$. We can then write:

$$M \models_{\mathcal{A}}^{P} N =_{def} P(M|N) \tag{19}$$

and construct (extended, probabilistic) logic infomorphisms as above, requiring that they preserve the conditional probabilities $P(M|N)$ for all subsets $M$, $N$ at each level of the hierarchy.[6] Traversing upwards in the hierarchy then imposes multiple conditioning on each "low-level" probability distribution; traversing downwards sequentially unpacks this conditioning. In fundamental Bayesian terms, $M$ above can be regarded as a previous event, whether observed or conjectured, and $N$ as a currently observed datum, in which

---

[6]As pointed out in [105], it is instructive to see that Eq. (19) reveals how a conditional probability can be used for interpreting the logical implication "$\Rightarrow$" as discussed in [108].

case $P(M)$ becomes the prior, and $P(N)$ the evidence, that together generate a prediction. Given the likelihood $P(N|M)$ as the conditional obtained from weakening the sequent via Eq. (19), Bayes' theorem specifies this conditional as the posterior:

$$P(M|N) = \frac{P(N|M)P(M)}{P(N)}. \tag{20}$$

A brief example illustrates these principles as follows: consider arbitrary classifiers $\mathcal{A}_1^{(a)}, \ldots, \mathcal{A}_5^{(e)}$ in some part of an information channel where (as in [99]) the classifiers correspond to events $a, b, c, d, e$, respectively, together with logic infomorphisms $f_{13}, \ldots, f_{45}$ between them, in which the sequents are relaxed to conditional probabilities via Eq. (19):



$$\tag{21}$$
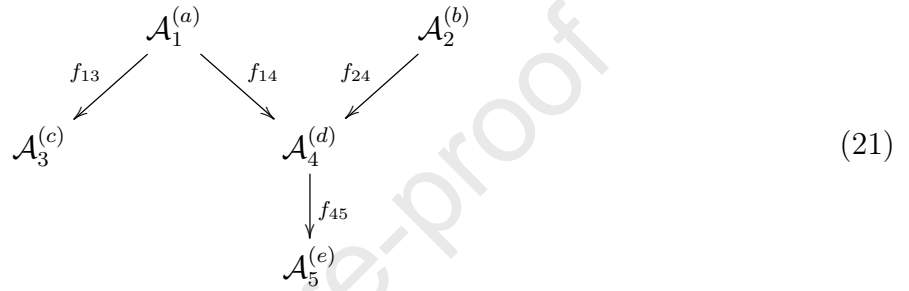
Following e.g. [104], this particular channel then generates a joint probability distribution given by:

$$p(abcde) = p(a)p(b)p(c|a)p(d|ab)p(e|d). \tag{22}$$

Putting these details within the framework of the above diagrams, a portion of a typical CCD computing a hierarchical Bayesian inference from a set of (posterior) observations $\mathcal{A}_i$ to an outcome $\mathbf{C}'$ has the form:



$$\tag{23}$$
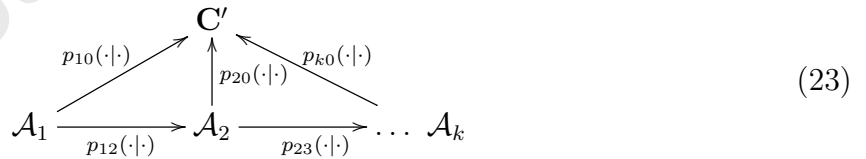
In this formal setting, the diagram commutativity of Diagrams (13) – (16), with probabilities added as in Diagram (23) above, enforces Bayesian coherence: the probability associated with any arrow $\mathcal{U} \to \mathcal{V}$ must equal the product of the probabilities associated with the arrows on any other directed path from $\mathcal{U}$ to $\mathcal{V}$.[7] Any pair of subsets of the $\mathcal{A}_i$ have, therefore, a well-defined joint probability distribution. As shown in [99], the converse is also true. Letting $\mathcal{A}_1, \mathcal{A}_2 \ldots \mathcal{B}$ and $\mathcal{B}, \mathcal{C}_1, \ldots \mathcal{C}_k$ be finite sets of classifiers, we can state the following [99, Thm. 7.1]:

---

[7]Probability spaces and conditional distributions can also be defined directly for Chu spaces as exhibited in [99, Exs. 2.4, 2.5]. See [106] for an alternative representation of Bayesian inference and derivation of the FEP using monodial categories and an operational formalism closely related to those employed in categorical quantum theory [107].

**Theorem 1.** *A well-defined joint probability distribution exists over subsets $\mathcal{A}_1, \mathcal{A}_2 \ldots \mathcal{B}$ and $\mathcal{B}, \mathcal{C}_1, \ldots \mathcal{C}_k$ of classifiers if and only if the following diagram commutes:*

$$
\begin{array}{c}
\mathbf{C} \\
\phi \nearrow \quad \nwarrow \psi \\
\mathbf{C_1} \qquad\qquad \mathbf{C_2} \\
f_1 \nearrow \quad \uparrow f_2 \quad \nwarrow f_B \quad g_B \nearrow \quad \uparrow g_2 \quad \nwarrow g_k \\
\mathcal{A}_1 \longrightarrow \mathcal{A}_2 \longrightarrow \ldots \mathcal{B} \longrightarrow \mathcal{C}_1 \longrightarrow \ldots \mathcal{C}_k
\end{array}
\tag{24}
$$

*that is, if and only if there exist logic infomorphisms $\phi$, $\psi$ and a classifier $\mathbf{C}$ such that the above diagram is a CCD.*

*Proof.* See [99]. The key observation is that $\phi$ and/or $\psi$ can only fail to exist if the joint probability distribution fails to exist. $\qquad\square$

The use of overlapping subsets of classifiers in Theorem 1 mirrors the use of mutually-noncommuting subsets of mutually-commuting observables in the canonical definition of quantum contextuality [53, 54] and its extension to a purely statistical criterion of incompatibility between measurement contexts [109, 110]. How the deployment of overlapping but mutually-noncommuting subsets of QRFs implements context-switching between pairs of complementary observables – such as position and momentum – is discussed in §3.4 below. How the FEP can drive context-switching as a component of active inference is then discussed in §4.3 below.

The various CCCD diagrams presented here in their hierarchical form, naturally represent both bottom-up and top-down causation, and thus lend themselves to informational descriptions of multiple realisations of higher level states by lower level states. Given suitable classifiers constituting a corresponding information flow, this is easily seen when "tokens" are simply the "realisations" of their corresponding "types". In fact, the core (colimit) $\mathbf{C}$ above, itself a classifier, encapsulates the total information flow of the CCCD; hence as a colimit, a"final realisation" (for discussions of multiple realisations in relationship to "agents", see e.g. [111, 112, 113])[8].

Viewing commutativity – and hence Bayesian coherence – as fundamental to the definition of measurement, and hence also to the definition of preparation or action on the environment generally, suggests that the Born rule, i.e. that if a state:

$$
|X\rangle = \sum_i \alpha_i |x_i\rangle,
\tag{25}
$$

the probability of obtaining $|x_i\rangle$ as an observational outcome is:

$$
P(|x_i\rangle = |\alpha_i|^2,
\tag{26}
$$

---

[8]The Map Example of [102, §2] nicely exemplifies the idea of "multiple realisation" in this context.

is a prescription for coherently assigning amplitudes $\alpha_i$ to components $|x_i\rangle$, consistent with the position advocated in [42, 44]. Note that as preparation and measurement are defined as operations of boundary states, any state $|X\rangle$ to which the Born rule applies must be on some boundary that functions as an MB.

# 3 Repeated measurements and system identification

## 3.1 Memory, time, and coarse-graining

The idea that a system must possess a (quasi-) NESS solution to its density dynamics – and hence be restricted to trajectories in its classical configuration space that do not diverge exponentially over time – in order to be observable as a "thing", immediately raises the issues of time as measurable duration and of memory as persistent over measurable time. As the simplest case, consider the state density $\rho_E$ of the observed environment sector $E$ of some agent $A$. The agent $A$ can only detect changes in $\rho_E$ and employ them as bases for inferences and actions if $A$ can write time-ordered records $[\rho_E(t_A)]$ to, and subsequently read them from, the memory sector $Y$. Here $t_A$ is an $A$-specific time coordinate that must be constructed. Persistence of the memory record $[\rho_E(t_A)]$ between writing and reading requires that it be "protected" from environmental noise, i.e. from ongoing events in $E$; hence if $|Y\rangle$ is the state encoding $[\rho_E(t_A)]$, $|Y\rangle$ must vary only slowly under the action of $H_B$. In the notation of Eq. (6), "protection" occurs if the coefficients $\alpha_i^B$ of operators $M_i^B$ acting on qubits within $Y$ all have small magnitudes. We can, therefore, distinguish two classes of actions by $A$ on its boundary/MB $\mathscr{B}$. Actions on $E$ are "questions" in Wheeler's [72] sense; they provoke informative responses from $B$ that can be used to develop a generative model of $H_B$ as discussed below. Actions on $Y$ are recordings that must remain relatively stable to be useful. Stability of the memory sector $Y$ against $H_B$ is a critical resource for any system $A$ capable of responding to environmental state changes, and hence of any $A$ capable of active inference. Heuristically, this is clearly evinced in active vision, where we actively palpate the world, every 250 ms or so, to update our working memory $Y$; i.e., update our (Bayesian) beliefs about the observed environment $E$.

The Second Law tells us that protecting any state against noise requires the expenditure of free energy. Hence, any agent $A$ is faced with a fundamental thermodynamic tradeoff: maintaining the stability of $Y$ requires free energy sourced from the unobserved (traced over) environmental sector $F$. For fixed $H_{AB}$, and hence fixed $\mathscr{B}$, expanding $F$ to obtain more free energy for $Y$ requires shrinking $E$, i.e. "seeing" less of the environment. The alternative is to coarse-grain $Y$ either in time or in space, i.e. in bit-string length. Which horn of this thermodynamic trilemma a given system takes is determined by the QRFs it implements. Classical manifestation of this trade-off are seen in many guises. For example, the intimate relationship between the efficiency of information transfer afforded by compression – as seen in minimum message length formulations of variational free energy [114, 115] – through to the minimisation of statistical complexity afforded by coarse-graining and quantisation [116].

The simplest memory-write process is illustrated in cartoon form in Fig. 3. A state $|E\rangle$, a bit string of length $\dim(E)$, is constructed from one-bit operators $M_i^E$ by a QRF $\mathbf{E}$. It is then written to the memory $Y$ by a QRF $\mathbf{Y}$, with free energy sourced from the remainder of $\mathscr{B}$, i.e. from the unobserved sector $F$. In the simplest case, the memory capacity $\dim(Y) = n\dim(E) + \log_2 n$ where $n$ is the number of distinguishable records. The $\log_2 n$ labels that allow records to be distinguished can, without loss of generality, be considered to be an integer sequence of clock ticks $i \to i+1$, starting from $i = 1$. Hence, any memory with more than one-record capacity defines a clock $\mathcal{G}_{ij}$, which we will see below must, in general, be a groupoid operator [46].[9] This clock is an internal time QRF that defines the time coordinate $t_A$.
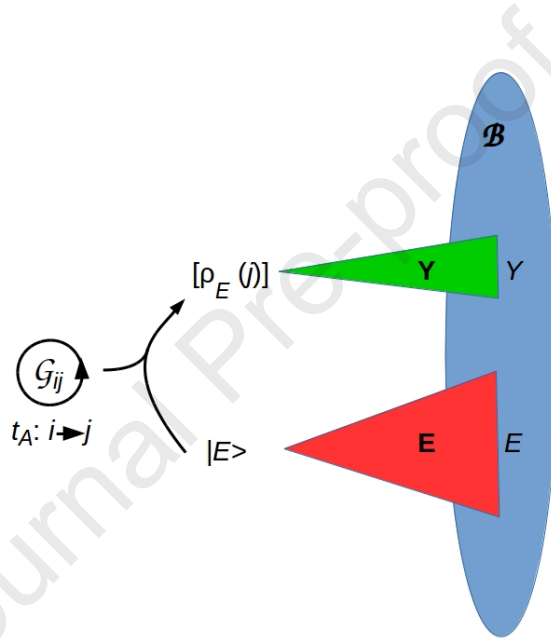


Figure 3: Cartoon illustration of QRFs required to write a readable memory of an observed environmental state $|E\rangle$. The QRFs $\mathbf{E}$ and $\mathbf{Y}$ read the state from $E$ and write it to $Y$ respectively. The clock $\mathcal{G}_{ij}$ is a time QRF that defines the time coordinate $t_A$.

The thermodynamic trade-off faced by $A$ is between $\dim(E)$, $\dim(Y)$, and the sampling time of states $|E\rangle$, which determines the length of one clock tick and hence of one unit of $t_A$. As mentioned earlier, the minimum timescale for biological systems is set by the thermal dissipation time to roughly 50 fs. Practical biological clocks run much more slowly; a clock

---

[9]Recall that a *groupoid* is a category in which the objects and arrows each comprise a set (as for a 'small' category), and every arrow is invertible [117, 118]. A groupoid generalizes the *group* concept in so far that the former can admit "multiple identities" (the objects).

based on Gamma-band neural activity in mammalian cortex, for example, has a sampling time of 10 – 20 ms. The recorded record, in this case, is a sample from or average over a time ensemble $<|E>>$ of measured states, i.e., is a record $[\rho_E]$ of a coarse-grained density. This record can be further compressed to achieve $\dim([\rho_E]) < \dim(E)$ and hence $\dim(Y) < n \cdot \dim(E) + \log_2 n$.

With these QRFs, the memory read-compare-write cycle can be represented as in Fig. 4. The classical record $[\rho_E(i)]$ written in the previous cycle is read by $\mathbf{Y}$ at "external" (i.e. Eq. (1) or (7)) time $t$. It is compared to the current measured state $|E(t)\rangle$ and a new record $[\rho_E(j)]$ is written by $\mathbf{Y}$ at $t+\Delta t$. This read-compare-write cycle advances the internal clock $\mathcal{G}_{ij}$ by one tick $i \to j$. All QRFs together with the comparison function are implemented by the internal dynamics $\mathcal{P}_A$. Formally, we can think of $\mathcal{P}_A$ as a weighted sum of "all possible paths" from the boundary state $|\mathscr{B}(t)\rangle$ to the boundary state $|\mathscr{B}(t + \Delta t)\rangle$ as in Eq. (8) [119]; see [120] for discussion.
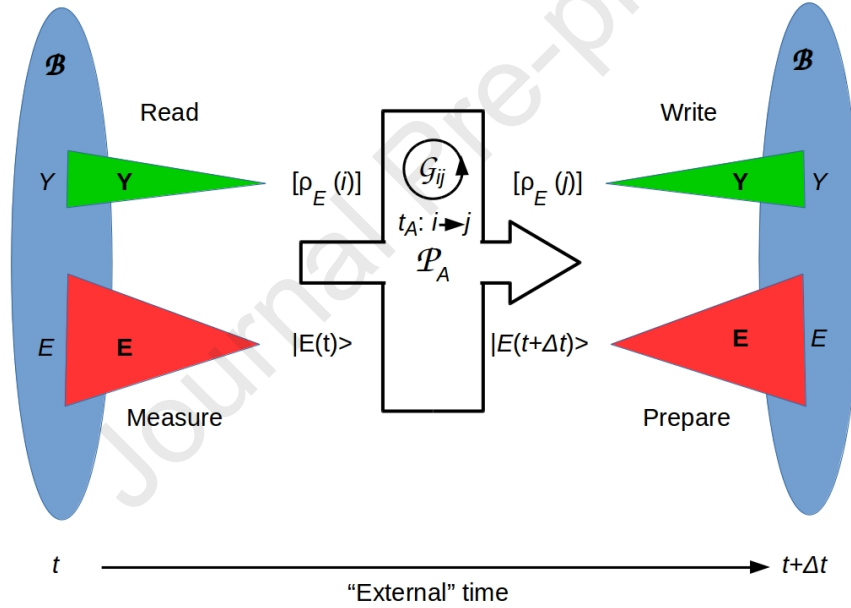


Figure 4: Cartoon illustration of one memory read-compare-write cycle defining one tick $i \to j$ of the internal clock $\mathcal{G}_{ij}$ and requiring an interval $\Delta t$ of "external" time $t$. All QRFs and the comparison function are implemented by the quantum dynamics $\mathcal{P}_A$ (block arrow).

Fig. 4 explicitly illustrates an important distinction between classical and quantum representations of dynamics and hence between classical and quantum formulations of the FEP. Classical physics assumes a spacetime embedding; hence the MB of any topologically-connected system can be associated with a spatial boundary that follows a smooth space-

25

time trajectory. "Internal" states of the system and the "internal" dynamics that operates on those states to maintain a NESS are spatially localized inside this boundary, and are, in particular, not exposed *on* the boundary. The current, quantum setting makes, in contrast, no assumptions about spacetime embedding as emphasized earlier. The boundary $\mathscr{B}$ represents a Hilbert-space decomposition, not a "physical" spacetime decomposition. As $\mathscr{B}$ is the only locus of classical information, "internal" classical states, including all classical memory records, are exposed on $\mathscr{B}$. The assumption that these exposed states are "protected" from $H_B$ corresponds to the classical assumption of "self-evidencing": that maintaining the integrity of the MB as a spatial boundary enables the maintenance of the NESS and vice-versa [10].

The true internal state of a quantum system $A$ is the state $|A\rangle$ that remains independently well-defined provided the joint state $|AB\rangle$ remains separable, i.e. provided the interaction $H_{AB}$ can be written as Eq. (6). This internal state is "protected" from $H_B$ by definition. Hence for quantum systems, "self-evidencing" is logically equivalent to separability. Classical memories are at risk of environmental perturbation, unless sufficient free energy is devoted to maintaining them. Quantum "memories" are encoded by the dynamics $\mathcal{P}_A$ and are at risk of environmental perturbation only if separability breaks down, i.e. if $|AB\rangle$ becomes entangled.

With this non-spatial understanding of the "internal" states of a quantum system $A$ with a classical memory, both the quantum $|A(t)\rangle \to |A(t+\Delta t)\rangle$ and the classical $[\rho_E(i)] \to [\rho_E(j)]$ loops are internal to $A$. Hence, the (classical) integrated information $\Phi(A) > 0$ as defined in Integrated Information Theory (IIT) [121], rendering $A$ "conscious" and hence an "agent" in the framework of IIT. Hence the notion of agency in IIT is consistent with that of Definition 1.

## 3.2 Learning and generative models

Learning a generative model allowing predictions of the future behavior of $E$ is straightforward in this setting. The prediction problem can be represented as follows:

$$\underbrace{[\rho_E(1)], [\rho_E(2)], \ldots [\rho_E(k-1)]}_{Prior}, \underbrace{[\rho_E(k)]}_{Posterior} \to \underbrace{[\rho_E(k+1)]}_{Prediction}, \tag{27}$$

where $[\rho_E(1)], [\rho_E(2)], \ldots [\rho_E(k-1)]$ are previous memory records, $[\rho_E(k)]$ is the current record, and $[\rho_E(k+1)]$ is the not-yet-obtained next record. The prior data can be re-represented in three progressively more sophisticated ways:

1. As a probability distribution over the set of possible records, a discrete set of no more than $2^{\dim(E)}$ elements.

2. As a probability distribution over the set of pairs $([\rho_E(i)], [\rho_E(i+1)])$ and hence as a discrete Markov process.

3. As a map from probability distributions on records $1, 2, \ldots i$ to probability distributions on records $1, 2, \ldots i + 1$ and hence as a discrete Markov kernel.

Representing the prior data by a discrete Markov kernel provides the greatest data compression, at the cost of more sophisticated processing by $\mathcal{P}_A$.

The process of updating a Markov kernel $\mathbb{M}_E^A(i)$ representing $A$'s prior data for $E$ can be formalized as:

$$\mathscr{L} : (\mathbb{M}_E^A(i), \ [\rho_E(i)]) \mapsto \mathbb{M}_E^A(i + 1), \tag{28}$$

where $\mathscr{L}$ is an operator of the form $(function, data) \to function'$, i.e. a learning operator. Hence any system $A$ that stores prior information using a data structure more space-efficient than an explicit linked list can be considered to be learning.

As the records $[\rho_E(i)]$ are just bit strings and Bayesian coherence is guaranteed by the commutativity of the operators composing the QRF $\mathbf{Y}$, the sequence $[\rho_E(1)], [\rho_E(2)], \ldots [\rho_E(k)]$ can be represented by a "true" Markov kernel $\mathbb{M}_E(k)$ satisfying the following commutativity constraint:

$$\begin{array}{ccc} \rho_E(i) & \xrightarrow{\ \mathbb{M}_E\ } & \rho_E(i + 1) \\ \mathbf{E} \Big\uparrow & & \Big\uparrow \mathbf{E} \\ |\mathscr{B}(t)\rangle & \xrightarrow{\ \mathcal{P}_U\ } & |\mathscr{B}(t + \Delta t)\rangle \end{array} \tag{29}$$

where as before one "tick" of $t_A$ corresponds to $\Delta t$ externally and the notation $|\mathscr{B}\rangle$ indicates the state of the qubit array encoded on $\mathscr{B}$. At step $k$ in $A$'s acquisition of state information about $E$, therefore, $A$'s prediction error $Er_E$ for $E$ is:

$$Er_E(k) = d(\mathbb{M}_E^A(k), \mathbb{M}_E(k)), \tag{30}$$

where the $d$ is the metric distance on Markov kernels. This definition is independent of $\mathscr{L}$ and hence of the sources of $A$'s prediction errors.

## 3.3 Identifying and measuring systems embedded in $E$

We have so far considered only observers $A$ that measure the states of their observed environments $E$ without decomposing $E$ into "systems" that have their own specific states. Such undifferentiated measurements of $E$ plausibly characterize all biological systems that interact with their environments primarily biochemically, instead of mechanically [122]. A bacterium measuring an ambient salt concentration, for example, does not assign the concentration value to a specific object within the environment [13]. Animals as diverse as arthropods, cephalopods, and vertebrates, however, detect and track the states of specific

external "particles" or objects, often other animals. In terms of theoretical biology, this ability is either ancient, arising at least by the Cambrian explosion, or results from convergent evolution in multiple distinct lineages.

The question of how an observer $A$ distinguishes a system $S$ from the environment $\tilde{E}$, in which $S$ is embedded, is central to classical cybernetics [20, 21] and, under the rubric of object persistence, to cognitive and developmental psychology [123, 124]. Here, $\tilde{E}$ indicates the remainder of $E$ when $S$ is removed, i.e. $E = S\tilde{E}$. While the question of how an observer distinguishes an external system from its surrounding environment, prior to – and as a precondition for – measuring some state of interest, is often neglected by physicists, it imposes significant thermodynamic and computational requirements on observers [125]. Distinguishing a system from its environment requires measurement, so it is naturally formulated in the language of QRFs.

To set up some notation, we will consider any distinct, identified system $S$ to comprise two components, $S = PR$, where $P$ is the "pointer" component that indicates some state $|P\rangle$ (or density $\rho_P$ of time-averaged samples of $|P\rangle$) of interest and $R$ is a "reference" component that by maintaining a constant state $|R\rangle$ (or constant density $\rho_R$ of time-averaged samples of $|R\rangle$) allows $S$ to be re-identified while $|P\rangle$ varies. Identifying a laboratory apparatus by monitoring a time-invariant reference state provides an example; see Fig. 5.
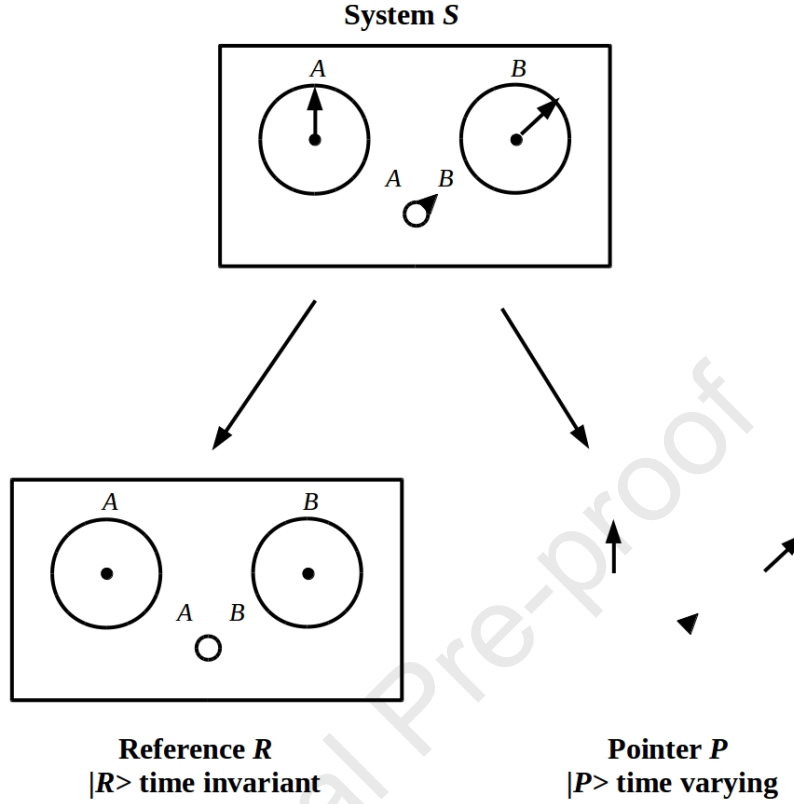
Figure 5: Identifying a system $S$ requires identifying some proper component $R$ that maintains a constant state $|R\rangle$ (or density of time-averaged samples $\rho_R$) as the "pointer" state $|P\rangle$ (or density of time-averaged samples $\rho_P$) of interest varies. Adapted from [46] Fig. 2, CC-BY license.

Following the reasoning above, decomposing $E$ into disjoint components $E = PR\tilde{E}$ is defining subsets of operators $\{M_i^E\} = \{M_j^P\} \sqcup \{M_k^R\} \sqcup \{M_l^{\tilde{E}}\}$ where $\sqcup$ indicates disjoint union. We can then define several QRFs:

$$
\begin{aligned}
\mathbf{P} &: P = \text{dom}(\{M_j^P\}) \to |P\rangle, \\
\mathbf{R} &: R = \text{dom}(\{M_k^R\}) \to |R\rangle, \\
\tilde{\mathbf{E}} &: \tilde{E} = \text{dom}(\{M_l^{\tilde{E}}\}) \to |\tilde{E}\rangle,
\end{aligned}
\tag{31}
$$

with corresponding memory records $[\rho_P(i)]$, $[\rho_R(i)]$, and $[\rho_{\tilde{E}}(i)]$. As QRFs that prepare as well as measure their assigned sectors of $\mathscr{B}$, these QRFs $\mathbf{P}$, $\mathbf{R}$, and $\tilde{\mathbf{E}}$ can be represented as

CCCDs with the symmetric form of Diagram (15). The processing pathway from a given sector $X$ to its associated memory record $[\rho_X]$ can, assuming a coarse-grained memory, be represented as a CCCD with the asymmetric form of Diagram (16); the memory-read to preparation of $X$ pathway has the opposite asymmetry. Note that identifiability of $S$ requires $[\rho_R(i)] = [\rho_R(j)]$ for all $i, j$. Forgetting what someone looks like, for example, can lead to re-identification failure.

Measurements of $P$, $R$, and $\tilde{E}$ are of no use unless they are recorded. As above, Markov kernels provide the most efficient data structure. The kernel $\mathbb{M}_R^A$ must be constant to enable system identification. By analogy with Eq. (28), these kernels can be associated with a learning operators $\mathscr{L}_P$, $\mathscr{L}_R$, and $\mathscr{L}_{\tilde{E}}$; where there is no requirement that these employ the same learning algorithm. The "true" Markov kernel $\mathbb{M}_E(i)$ can similarly be decomposed into components, each of which must satisfy the commutativity constraint expressed by Diagram (29):

$$
\begin{array}{ccc}
\rho_P \xrightarrow{\mathbb{M}_P(i)} \rho_P(i+1) & \rho_R(i) \xrightarrow{\mathbb{M}_R} \rho_R(i+1) & \rho_{\tilde{E}}(i) \xrightarrow{\mathbb{M}_E} \rho_{\tilde{E}}(i+1) \\[4pt]
\mathbf{P}\uparrow \qquad \uparrow\mathbf{P} & \mathbf{R}\uparrow \qquad \uparrow\mathbf{R} & \tilde{\mathbf{E}}\uparrow \qquad \uparrow\tilde{\mathbf{E}} \\[4pt]
|\mathscr{B}(t)\rangle \xrightarrow{\mathcal{P}_U} |\mathscr{B}(t+\Delta t)\rangle & |\mathscr{B}(t)\rangle \xrightarrow{\mathcal{P}_U} |\mathscr{B}(t+\Delta t)\rangle & |\mathscr{B}(t)\rangle \xrightarrow{\mathcal{P}_U} |\mathscr{B}(t+\Delta t)\rangle
\end{array}
\tag{32}
$$

Hence prediction errors for $P$, $R$, and $\tilde{E}$ can be defined as in Eq. (30). Failing to correctly predict $\rho_R$ results in system-identification failure and renders concurrent observations of $\rho_P$ meaningless.

The "systems" $P$, $R$, and $\tilde{E}$ are, clearly, just subsets of outcomes obtained by measuring the qubits on $A$'s boundary/MB $\mathscr{B}$; their states are, therefore, determined by the actions of $B$'s encoding operators $M_i^B$. As $[M_i^B, M_j^B] = 0$ for all $i, j$ by definition, $R$, $P$, and $\tilde{E}$ must all be mutually separable and hence mutually decoherent; equivalently, $\mathbf{R}$, $\mathbf{P}$, and $\tilde{\mathbf{E}}$ must all mutually commute. Hence, $A$ can regard $R$, $P$, and $\tilde{E}$ as "things" with distinct identities and states as required by [10]. It bears emphasis, however, that no spacetime background has been assumed in writing Eq. (6), in defining $\mathscr{B}$, or in defining any of the sectors $R$, $P$, or $\tilde{E}$. The "things" $R$, $P$, and $\tilde{E}$ are not, therefore, observer-independent in any sense, although observers deploying similar QRFs and able to communicate (i.e. deploying QRFs that enable mutual recognition and communication) may agree as to their states [44, 45]. The present formalism is, therefore, "system-free" in the sense of [43] and hence represents measurements using external apparatus as fully device-independent. While device independence is implied whenever an MB is invoked [11], classical formulations of measurement interactions – indeed, of perception generally – tend nonetheless to assume differentiated external objects *a priori*, as Einstein's famous insistence that the Moon is there when no one is looking exemplifies [126]; see [127, 128, 129] for examples from perceptual psychophysics. Evolutionary considerations argue against any assumption of *a priori* objects or even Galilean invariances of motion [130, 131], consistent with the background-independent

approach taken here.

## 3.4   Noncommutativity and context-switching

As emphasized by Bohr almost 100 years ago [132], a finite quantum of action $\hbar$ partitions the set of all possible quantum measurement operators into a set of "complementary" non-commuting pairs, the most well-known being position $\hat{x}$ and momentum $\hat{p}$. These operators, as well as all other operators acting on "systems" with associated spacetime coordinates, correspond in the current background-free framework to QRFs acting on sectors of $\mathscr{B}$, i.e., on subsets of qubits as described above. Hence, noncommuting operators correspond to noncommuting QRFs, as formalized by Theorem 1.

Two QRFs $\mathbf{U}$ and $\mathbf{V}$ can fail to commute only if the underlying measurement operators fail to commute. However, as noted previously, any set of operators $M_i^k$ appearing in Eq. (6) must all mutually commute. Switching between noncommuting QRFs $\mathbf{U}$ and $\mathbf{V}$, therefore, entails switching between a mutually-commuting operator set $M_i^A$, of which the $M_j^U$ are a subset, and a complementary mutually-commuting operator set $O_i^A$, of which the $O_j^V$ are a subset, where for at least some $i, j$, $[M_i^U, O_j^V] \neq 0$. This switch implements a basis rotation on $H_{AB}$, leaving its dimension $N$ and its eigenvalues, the binary representations of which are the bit strings encodable on $\mathscr{B}$, both unchanged while replacing the amplitudes $\alpha_i^A$ with amplitudes $\lambda_i^A$ so that Eq. (6) now reads, for $A$:

$$H_{AB} = \beta^A k_B T^A \sum_i^N \lambda_i^A O_i^A. \tag{33}$$

In practice, we will be primarily interested in partial basis rotations in which the $M_i^A$ and the $O_i^A$ substantially overlap, e.g. maintaining fixed sectors $F$, $\tilde{E}$ and $R$ while switching between complementary pointer sectors as discussed below. Note that such basis rotations have no effect on $B$ or its operators, so are undetectable, in principle, by $B$.

An observer $A$ capable of switching between noncommuting QRFs must, to maintain an operable memory, implement a clock that is invariant under basis rotations on $H_{AB}$. If measurements made at clock ticks $i$ and $j$ do not commute, however, the corresponding clock operations will not commute; in particular $\mathcal{G}_{ij} \circ \mathcal{G}_{ji} \neq \mathcal{G}_{ji} \circ \mathcal{G}_{ij}$ where $\circ$ is operator composition. It is for this reason that the $\mathcal{G}_{ij}$ form a groupoid, and not a group [46]. From a more practical perspective, noncommutativity forces $t_A$ to be unidirectional, and hence memory records to be encoded irreversibly with an accompanying expenditure of free energy. The internal clock $\mathcal{G}_{ij}$ thus defines $t_A$ as an entropic time, consistent with the analysis in [133]. Any observer $A$, therefore, observes a unidirectional flow of information from $B$ and of dissipated heat to $B$; hence any observer $A$ confirms the Second Law with respect to its internal time $t_A$. As $A$ and $B$ are completely symmetric by Eq. (6), $B$ also confirms the Second Law with respect to $t_B$, showing that the Second Law is observer-relative and independent of the "external" time $t$, consistent with the analysis in [134].

Switching between noncommuting QRFs while holding other QRFs constant, e.g., switching between interference (position) and which-path (momentum) measurements on an interferometer identified by a fixed reference sector $R$, is switching between mutually-noncommuting but overlapping sets of mutually-commuting operators as described in Theorem 1 and is a canonical quantum context switch [53, 54]. Pointer-state observables are, in particular, *observables in context* as defined in [99]: the state $|P\rangle$ of any pointer sector $P$ is measured in the context of the separable joint state $|R\rangle|\tilde{E}\rangle|F\rangle$, where the component $|F\rangle$ is unobserved by definition. By Theorem 1, two sets of pointer operators $M_j^U$ and $O_j^V$ as above, that define alternative pointer sectors $U$ and $V$ by Eq. (10), are mutually-commuting and hence *co-deployable* if and only if maps $\phi$ and $\psi$ exist such that the diagram:

$$
\begin{array}{ccccc}
 & & \mathbf{C} & & \\
 & {}^{\phi}\nearrow & & \nwarrow^{\psi} & \\
\mathbf{C_1} & & & & \mathbf{C_2} \\
{}^{f_1}\uparrow \quad {}_{g_1}\nearrow & & & \nwarrow_{g_2} \quad \uparrow^{f_2} & \\
U & \longrightarrow & R\tilde{E}F & \longrightarrow & V
\end{array}
\tag{34}
$$

commutes. Here, we have replaced the explicit operators in Diagram (24) by their corresponding sectors to simplify the notation. We can equally well interpret $UF$ and $VF$ as contexts for the observation of sectors $R$ and $\tilde{E}$: in this case switching between $U$ and $V$—with the unobserved sector $F$ held fixed—yields consistent probability distributions if and only if Diagram (34) commutes.

If Diagram (34) fails to commute, then pointer-state observables are said to be *non-co-deployable*. Non-commutativity of the CCD in (34) had been specified in [99, Th. 7.1] in terms of non-existence of a consistently definable joint probability distribution of conditionals, such as for Diagram (23). This non-co-deployability of observables thus amounts to occurrence of intrinsic (quantum) contextuality in relationship to (34).[10]

We will see in §4.2 below that context switching increases variational free energy (VFE) by generating Bayesian "prediction errors"; hence context-switching makes minimizing VFE and hence complying with the FEP more difficult. Deploying noncommuting QRFs against a fixed background can, however, lead to radically better generative models, as the history of technological applications of quantum theory attests. Hence, context-switching poses a fundamental challenge to any classical formulation of the FEP, and a fundamental explanandum for a quantum formulation.

---

[10]As recalled in e.g. [135, §3], 'non-commutativity' is at the very heart of contextuality, as first formulated by von Neumann in terms of non-commutativity of self-adjoint operators representing measurement, with the impossibility of simultaneously measuring the eigenvalues corresponding to non-commuting operators. In summarizing the principal results of ensuing hidden variables theory, Mermin [136] demonstrated the impossibility of finding a joint probability distribution for all possible observables.

# 4 FEP for generic quantum systems

## 4.1 Defining VFE for quantum systems

The FEP is a variational or least-action principle: it states that a system enclosed by an MB, and therefore having internal states that are conditionally independent of its environment, will evolve in a way that tends to minimize a VFE that is an upper bound on surprisal. Formally, the VFE $F(\pi)$, where $\pi$ is a "particular" state $\pi = (b, \mu)$ comprising MB ($b$) and internal ($\mu$) components, can be written [10, Eq. 8.4],

$$F(\pi) \triangleq \underbrace{\mathfrak{I}(\pi)}_{Surprisal} + \underbrace{D_{KL}[Q_\mu(\eta) \parallel P(\eta|b)]}_{Divergence} \geq \mathfrak{I}(\pi). \tag{35}$$

This functional as an upper bound on surprisal $\mathfrak{I}(\pi) = -\log P(\pi)$ because the Kullback-Leibler divergence ($D_{KL}$) term is always non-negative. This KL divergence is between the density over external states $\eta$, given the MB state $b$, and a variational density $Q_\mu(\eta)$ over external states parameterised by the internal state $\mu$. If we view the internal state $\mu$ as encoding a posterior over the external state $\eta$, minimizing VFE is, effectively, minimizing a prediction error, under a generative model supplied by the NESS density. In this treatment, the NESS density becomes a probabilistic specification of the relationship between external or environmental states and particular (i.e. "self") states.

In the notation developed in §2 and 3 above, we can write the surprisal for a quantum system $A$ in its most general form as:

$$\mathfrak{I}^A(t) = -P(|\mathscr{B}(t)\rangle \mid |A(t)\rangle) \tag{36}$$

and the corresponding evidence bound as:

$$D_{KL}[Q_{|\mathscr{B}(t)\rangle}(|B(t)\rangle) \parallel P(|B(t)\rangle \mid |\mathscr{B}(t)\rangle)]. \tag{37}$$

In the current setting, however, these expressions have little direct utility, as our effective starting point, Eq. (6), constrains neither $|A(t)\rangle$ nor $|B(t)\rangle$. Indeed, from a strict formal perspective, neither Eq. (36) nor Eq. (37) is well-defined in the current setting. The full boundary/MB state $|\mathscr{B}(t)\rangle$ is, moreover, not an observable for $A$ (or $B$), as the thermodynamic-resource sector $F$ remains unobservable by definition. We have, however, already derived in §3.2 a representation of $A$'s prediction error, Eq. (30), which we reproduce here for reference:

$$Er_E(k) = d(\mathbb{M}_E^A(k), \mathbb{M}_E(k)). \tag{38}$$

In this expression, the timestep $k$ counts $A$'s internal clock time $t_A$ and the kernels $\mathbb{M}_E^A$ and $\mathbb{M}_E$ are derived from observables and therefore constrained by the theory. The kernel $\mathbb{M}_E(k)$ represents the observable behavior of $\mathscr{B}$, as localized to the sector $E$, up to $t_A = k$.

33

The kernel $\mathbb{M}_E^A(k)$ is $A$'s generative model of the action of the unknown dynamics $\mathcal{P}^B(t)$ on $\mathscr{B}$, also as localized to $E$. Hence $Er_E(k)$ represents $A$'s total *reducible* uncertainty about $B$ at $t_A = k$. It is, therefore, an upper bound on surprisal analogous, in the current setting, to $F(\pi)$.

The operators $M_i^E$ referenced by Eq. (30), i.e. (38) must, clearly, all be co-deployable. In practice, however, $E$ is as discussed above subject to context switches of the form $UR\tilde{E} \to VR\tilde{E}$ whenever $A$ switches between noncommuting pointer QRFs $\mathbf{U}$ and $\mathbf{V}$ and hence non-co-deployable operators $M_i^U$ and $M_i^V$. Hence $Er_E$ is only well-defined in the absence of context switches; in the presence of context switches the generalized uncertainty relation:

$$\Delta u \Delta v \geq \hbar/2 \tag{39}$$

for pointer outcomes $u$ and $v$ can generate divergent uncertainties. Hence in practice, any system $A$ is faced with separately minimizing:

$$Er_X(k) = d(\mathbb{M}_X^A(k), \mathbb{M}_X(k)). \tag{40}$$

for each sector $X$ defined by a QRF $\mathbf{X}$. We can, therefore, formulate the FEP for generic quantum systems, taking context-switching into account, as:

> **FEP**: *A generic quantum system $A$ will act so as to minimize $Er_X$ for each deployable QRF $\mathbf{X}$.*

A trivial agent can be viewed as executing a trivial QRF, i.e. as only exercising choice of basis for writing to and reading from $\mathscr{B}$ as a whole, and so satisfies the FEP trivially.

## 4.2 Sources of VFE for quantum systems

As noted in the Introduction, there is no source of objective randomness in the current formalism. Indeed, an observer $A$ can be regarded as *certain* of the states $|\tilde{E}\rangle$, $|R\rangle$, $|P\rangle$, and $|Y\rangle$ of the observable sectors of $\mathscr{B}$ at every (external) time $t$. Uncertainty and prediction error – and hence, VFE – is generated in the current formalism by $A$'s in-principle ignorance of both the state $|B\rangle$ and the dynamics $\mathcal{P}_B$ of its interaction partner $B$. As the bits $A$ reads from $\mathscr{B}$ are written by $\mathcal{P}_B$, $A$'s ability to predict the future states of its observable sectors, and hence to minimize $Er_X$ for each sector $X$ via Eq. (40), depends on its ability to predict the behavior of $\mathcal{P}_B$ locally on each observable sector. As the thermodynamic sector $F$ is not observed, direct predictions on $F$ are not possible; the local behavior of $\mathcal{P}_B$ on $F$ can at best be predicted from its local behavior elsewhere. An animal, for example, must employ its available senses – hence its observable sectors – to predict the nutritional value of food.

The option space governing $A$'s ability to locally predict $\mathcal{P}_B$ is summarized in Fig. 6. What is important for $A$ is not the dynamic complexity or even the dimension of $\mathcal{P}_B$, both of which are unobservable in principle, but rather the dynamic complexity of the action of

$\mathcal{P}_B$ on $\mathscr{B}$ (the dimension of this action is, clearly, just the dimension of $\mathscr{B}$). Here, the weak-interaction limit that allows separability between $A$ and $B$ is significant: $H_{AB}$ (and hence $\mathscr{B}$) must have significantly lower dimension that $H_B$ (and hence $\mathcal{P}_B$) if the weak interaction limit to is hold. The simplest case is shown in Fig. 6, Panel a), in which the system $B$ is a trivial agent deploying no QRFs other that the choice of basis for interactions with $\mathscr{B}$. The action of $\mathcal{P}_B$ is, in this case, limited to choice of basis, e.g., to rotating the $z$ axis $z_B$ in Fig. 1. As discussed in §2.4, basis rotation by $B$ generates quantum noise in the communication channel defined by $H_{AB}$ that is indistinguishable by $A$ from classical noise. Hence, the trivial agent $B$ in Fig. 6, Panel a) "looks like" a noise source to $A$. Emission of Hawking radiation from a black hole (BH) provides perhaps the most pure example of such a noise source; while the dimension "inside" the BH can be arbitrarily large (see e.g. [76, Fig. 19]), the internal dynamics are uncoupled from the classical information encoded on the horizon and hence have no classical computational power. As will be discussed in §4.3 below, a "small" trivial agent will be driven by the FEP toward entanglement with the larger system $A$.
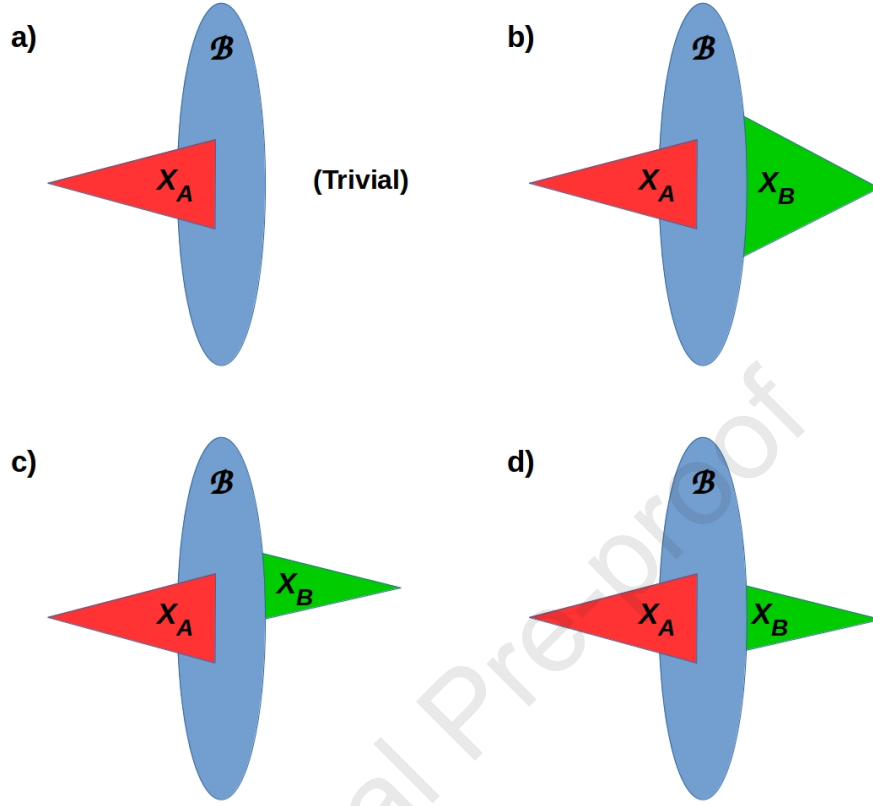
Figure 6: Four options for $A$'s ability to predict the local behavior of $\mathcal{P}_B$ on an observable sector $X_A$. a) A trivial agent deploying no QRFs beyond choice of basis for interacting with $\mathscr{B}$ appears as a noise source to $A$. b) $B$ encodes a sector $X_B$ that contains $X_A$; the bits on $X_B$ but outside $X_A$ encode "nonlocal hidden variables" for $A$. c) The sectors $X_A$ and $X_B$ overlap; the areas of non-overlap become noise sources. d) If $X_A = X_B$, VFE is generated by insufficient learning.

The more interesting parts of $A$'s option space for prediction are shown in Fig. 6, Panel b), c), and d), in which $B$ is nontrivial. If $B$ is nontrivial, it deploys at least one QRF $\mathbf{X_B}$ acting on a sector $X_B$. As discussed in §2.5, $B$'s sectors must be mutually decoherent, so the action of $\mathcal{P}_B$ on $X_B$ is independent of its action elsewhere; it is this independence that makes prediction possible. If $X_B$ does not overlap any *observable* sector for $A$, however, $B$ will appear trivial, i.e. as a noise source, to $A$. Hence, the interesting cases are the ones in which $A$'s and $B$'s observable sectors overlap; this is the case, intuitively, in which $A$ and $B$ can "see each other" and hence interact in the ordinary, nontechnical sense of that term.

In Fig. 6, Panel b), $B$'s sector $X_B$ fully contains $X_A$. As noted above in connection with Eq. (31), all measure – prepare QRFs $\mathbf{X}$ are symmetric, i.e $\mathrm{codom}(\mathbf{X}) = \mathrm{dom}(\mathbf{X})$. Preparation

by $B$ of the bits in $X_A$ will, therefore, in general depend on bits outside of $X_A$ but within $X_B$, i.e. on bits with the remainder $X_B \setminus X_A$. The values of these bits are "nonlocal hidden variables" [54] from $A$'s perspective; they affect what is observed on sector $X_A$ without being local to, i.e., contained within, $X_A$. Indeed, such bits may be within $F$ and hence unobservable in principle by $A$. Changes in the values of these nonlocal hidden variables are, effectively, context changes as defined in [109, 110]; probability distributions $P(X_A|\zeta)$ and $P(X_A|\xi)$ for distinct hidden-variable states $\zeta$ and $\xi$ may be different. The "context-blind" distribution $P(X_A)$ can, in this case, fail to be well-defined over time. Such failures manifest as violations of Leggett-Garg inequalities [57], i.e. as "quantum hysteresis" effects due to nonlocal (i.e. outside of $X_A$) and possibly unobservable causes. In the language of artificial intelligence or robotics, they appear as failures to solve the Frame Problem [99], the problem of predicting what will not change as the result of an action [137]. If $X_A$ is a reference sector for $A$, Frame Problem solution failures on $X_A$ can result in failures of object re-identification [138].

A situation in which $X_A$ fully contains $X_B$ presents similar issues to that in Fig. 6, Panel b), except here the "hidden variables" are in $X_A \setminus X_B$ and hence are accessible to $A$. The bits in $X_A \setminus X_B$ nonetheless contribute VFE – effectively, noise – to $X_A$ that is unconstrained by $X_B$. If $X_A$ and $X_B$ overlap with remainders, as in Fig. 6, Panel c), a similar noise contribution to $X_A$ (or on $B$'s side, to $X_B$) results. The final possibility is, clearly, that in which $X_A = X_B$ as shown in Fig. 6, Panel d). Here, the source of VFE is not noise, but rather differences in the computations implemented by the QRFs $\mathbf{X_A}$ and $\mathbf{X_B}$. Such differences correspond, in the notation of §2.6, to differences in the structures of the CCCDs implementing $\mathbf{X_A}$ and $\mathbf{X_B}$, e.g. differences in the "connection weights" if these are thought of as ANNs or VAEs. They correspond, in other words, to learning failures, e.g. due to insufficient training-set representativeness, as Eq. (28) renders obvious. We can, therefore, represent the overall situation for any observer $A$ as:

$$\text{VFE} \ = \ \text{Noise} \ + \ \text{Insufficient Learning} \tag{41}$$

consistently with Eq. (35) above. Here "noise" includes VFE generated by unobserved context changes (Leggett-Garg violations or Frame Problem solution failures) as well as "classical" noise.

## 4.3 Asymptotic behavior of the FEP

Having seen how VFE is generated, we can now ask how it is minimized: how, in other words, an agent acts in accordance with the FEP. As discussed in the Introduction, the FEP in its classical form is effectively the statement that any system with sufficient stability to be a "thing" – i.e. a system with a [quasi-] NESS density – will act so as to preserve its "thingness" by maintaining the integrity of its MB and hence the integrity of its "self" as distinct from its surroundings. Comparing Eq. (35) and (41), it becomes clear what this amounts to: a system "self evidences" by behaving in a way that minimizes noise while improving learning. This, as is well known, induces a trade-off: learning requires seeking

37

uncertainty in order to minimize it. As emphasized in [139], successful learning requires a focus on learnable tasks and avoidance of unlearnable tasks. Hence, minimizing VFE is removing all removable noise. It does not, in practice, lead to perfect predictability of future MB states, but to best feasible predictability of future MB states. In classical FEP formulations, this becomes clearly evident in the form of a functional called expected free energy (EFE); namely, VFE expected under a posterior predictive density that is conditioned on action. In this setting, the most probable actions are those that minimise EFE and thereby resolve uncertainty or maximise information gain (a.k.a., intrinsic motivation or epistemic affordance). In short, novelty-seeking is an emergent property of any"thing", under the FEP.

We can, however, ask in the current framework how the FEP behaves asymptotically, i.e., what are the consequences for $A$ as, in the notation of Eq. (40), $Er_X(k) \to 0$ as $k \to \infty$ for all observable sectors $X$. This clearly involves implementing a learning operator $\mathscr{L}_X$ for each sector $X$ that is capable of asymptotically-perfect learning: let us assume this is the case. What remains given Eq. (41) is noise, including noise due to observed context switches. Only one mechanism for removing noise is available: that shown in Fig. 6, Panel d). Hence we can conclude:

*The FEP asymptotically drives alignment of QRFs across $\mathscr{B}$.*

It drives any observer $A$, in particular, to match any context switches by its interaction partner $B$ in order to maintain QRF alignment. Let us now consider, therefore, a situation in which *all* QRFs deployed by $A$ and $B$ are aligned as in Fig. 6 d), and in which *all* of $A$'s QRFs have learned the local behavior of $\mathcal{P}_B$ on their sectors perfectly. The local behavior of $\mathcal{P}_B$ on some shared sector $X$ is determined by $B$'s QRF $\mathbf{X_B}$. This QRF $\mathbf{X_B}$ is, however, a quantum computation; as such, it encodes nonfungible – not finitely classically encodable – information as shown in [49]. The future behavior of $\mathbf{X_B}$ can, therefore, only be perfectly predicted by $\mathbf{X_B}$ itself, that is:

$$Er_X \to 0 \;\Rightarrow\; \mathbf{X_A} = \mathbf{X_B} \tag{42}$$

If $A$ and $B$ are separable, the consequent in Eq. (42) violates the no-cloning theorem [140]: it demands that the internal quantum state $|B|_X(t)\rangle$, the time (external $t$) evolution of which implements $\mathbf{X_B}$, be replicated exactly in $A$. Hence, if Eq. (42) holds, $A$ and $B$ cannot be separable. Therefore we have:

*If $AB$ is isolated, the FEP asymptotically drives the joint state $|AB\rangle$ to entanglement.*

The claim that isolated systems are driven to entanglement is familiar from our initial discussion of bipartite interactions in §2.3: all isolated systems are driven to entanglement by unitary evolution. Separability is a special case, an approximation that holds only under conditions of weak interactions and short observation times. Hence, we can, finally, conclude:

*The FEP is, asymptotically, the Principle of Unitarity.*

The FEP is, in other words, asymptotically equivalent to the first axiom of quantum theory. Any quantum system, therefore, *must* behave in accord with the FEP; doing so is simply approaching entanglement with its environment as required by the Principle of Unitarity. The FEP is therefore, consistent with the results obtained in [10], a fundamental, generic physical principle. Conversely, the Principle of Unitarity – the principle that observable information is conserved – can now be seen as a fundamental principle of cognitive science. It is, specifically, a fundamental principle of the cognitive science of interacting, complementary systems.

As an example of the FEP in action, let us return to the situation in which a "small" trivial system $B$ interacts with a larger, nontrivial system $A$ considered above. The only freedom $B$ has in this case is freedom of basis choice for reading and writing from $\mathscr{B}$. From $B$'s perspective, any basis choice that is misaligned with $A$'s basis choice generates noise. The FEP will, therefore, drive $B$ to align its choice of basis – $z$ axis in Fig. 1 – with that of $A$. This basis alignment, however, leaves $B$ entangled with $A$. This is not surprising. When a photon, for example, interacts with an atom, it is completely absorbed and loses its identity as a "thing"; its state becomes irreversibly entangled with that of the atom with which it interacts.

We can develop this result more formally as follows, noticing that a preparation operation by $\mathbf{B_X}$ followed by a measurement operation by $\mathbf{A_X}$ can be described by a CCCD in the form of Diagram (14), with the classifiers $\mathcal{A}_i$ identified with the bits comprising sector $X$ that are prepared and then measured. The question of asymptotic behavior is then the question of whether this CCCD is symmetric, i.e. whether the cores $\mathbf{C}' = \mathbf{D}'$ in Diagram (14). We approach this by considering the Markov kernels $\mathbb{M}_X^A$ and $\mathbb{M}_X$, the difference between which defines the error $Er_X$ (via Eq. (40)) that the FEP asymptotically sends to zero.

As shown in [141, p.17], following [142, §4], every category $\mathfrak{C}$ whose arrows form a set $K$ embeds fully as a subcategory into **Chu**. There is, therefore, an induced mapping $\mathscr{F} : \mathfrak{C} \longrightarrow \mathbf{Chu}$ realizing $\mathscr{F}(\mathfrak{C})$ as an embedded subcategory of **Chu** that consists of some objects of the latter, together with all of the arrows between them (for background on such embeddings, see e.g. [50, 51]).

Now we recall the direct association between Markov kernels and conditional probabilities, and apply the above embedding to the category $\mathfrak{P}$ of conditional probabilities, following mainly [143, 144] (cf. [145, 146]). Objects of $\mathfrak{P}$ consist of countably generated measurable spaces $(\mathcal{X}, \Sigma_{\mathcal{X}})$, and arrows of $\mathfrak{P}$ between two such objects, having the form:

$$\mathbb{M} : (\mathcal{X}, \Sigma_{\mathcal{X}}) \longrightarrow (\mathcal{Y}, \Sigma_{\mathcal{Y}}). \tag{43}$$

These arrows are Markov kernels assigning to each $x \in \mathcal{X}$, and each measurable set $Q \in \Sigma_{\mathcal{Y}}$, the probability of $Q$ given $x$, denoted here by $\mathbb{M}(Q|x)$, whenever defined. We can also write this as $p_{\mathbb{M}}(Q, |x)$, the (regular) conditional probability determined by the arrow $\mathbb{M}$, 'regular'

in so far that $\mathbb{M}$ is conditioned on points rather than on measurable sets in $\Sigma_{\mathcal{X}}$. As arrows given by Eq. (43) specify a Markov process, they comprise a semigroup, and therefore a set. Hence, we obtain a full embedding $\mathfrak{P} \longrightarrow \mathbf{Chu}$, where the (arrow) set $K$ is identified with a set of conditional probabilities $K = \{p_{\mathbb{M}}(S, |x)\} \subseteq [0, 1]$.[11] Using the fact that $\mathbf{Chu}$ and the Channel Theory category $\mathbf{Chan}$ are isomorphic categories with respect to their respective objects and arrows, we can summarize as follows:

**Proposition 1.** *The category $\mathfrak{P}$ embeds fully into $\mathbf{Chan}$ with classification $\Vdash_{\mathcal{A}}$ realized by the conditional probability $p_{\mathbb{M}}(\cdot|\cdot)$ (the Chu space valuation/satisfaction relation) whenever this is defined.*

Without loss of generality, we can assume the CCD in Diagram (23) to be a diagram in this embedded subcategory (for a survey of probability spaces and Bayesian belief networks in terms of the categories $\mathbf{Chu}$ and $\mathbf{Chan}$, see [99, §2]).

We can now proceed to consider the asymptotic behavior of the FEP, in particular the conditions under which $|\mathbb{M}_X - \mathbb{M}_X^A| \to 0$ for an arbitrary system $A$ and sector $X$ (with QRF $\mathbf{X}$), in the setting where $A$ interacts with some $B$ and the joint system $AB$ is isolated. To make sense of the formal difference $\mathbb{M}_A - \mathbb{M}_B$, that is, to make sense of the difference between two arrows in $\mathfrak{P}$, we can use the metric distance on Markov kernels as in Eq. (30). This distance is then manifestly the difference between the conditional probabilities $p_{\mathbb{M}_A}(\cdot|\cdot) - p_{\mathbb{M}_B}(\cdot|\cdot)$, in the way the corresponding Markov kernels determine these as described above. As discussed previously, this is a difference in the (conditional) probabilities of observable behavior in sectors. Thus, $|\mathbb{M}_A - \mathbb{M}_B| \to 0$ is interpreted as the metric distance $d(\mathbb{M}_A, \mathbb{M}_B) \to 0$ in the asymptotic limit.

With the labeling by $A, B$, let us return to (23) which we adopt to provide two separate CCDs as specified by (23) denoted by $\mathrm{CCD_A}$ and $\mathrm{CCD_B}$. The asymptotic limit $\to 0$ of the metric distance of the kernels determining the difference of the conditionals (hence $\to 0$ in the asymptotic limit) provides the sense in which the diagrams $\mathrm{CCD_A}$ and $\mathrm{CCD_B}$, along with their respective colimits $\mathrm{colim}(\mathbf{C}'_A)$ and $\mathrm{colim}(\mathbf{C}'_B)$, can be identified in this limit (or in the notation of Diagram (14), the cores $\mathbf{C}'$ and $\mathbf{D}'$ as required above become identical).

# 5  Discussion

## 5.1  High-level overview

While much of the foregoing has been technical, it is conceptually straightforward. To briefly review, reformulating the FEP within quantum information theory allows us to drop the assumption of an observer-independent spacetime background characterized by (continuous) frames of reference that can be specified to infinite precision, and also to

---

[11]In a similar way, Abramsky in [147] shows that the Dirac-von Neumann formulation of quantum mechanics can be conveniently represented in the category $\mathbf{Chu}$.

drop the assumption of randomness. In the quantum formulation, the blanket states are implemented by a holographic screen separating the interacting systems $A$ and $B$. The screen is the (topological) locus of the interaction $H_{AB}$; "sensory" and "active" states of the classical MB become incoming and outgoing encodings of bits on the screen. The interaction is symmetrical across the screen: the reciprocal exchange between "internal" $A$ and "external" $B$ systems takes the form of answers and questions (formulated as Hermitian operators $M_i^k$), where questions correspond to classical action (mediated by active states) and answers correspond to classical perception (mediated by sensory states). There is, crucially, no assumption that $A$ and $B$ share preparation and measurement bases; basis mismatches between $A$ and $B$ generate quantum noise that is "perceived" as classical noise.

An "agent" in the quantum formalism is a system that deploys quantum reference frames – effectively, *concepts* that identify persistent objects and their time-varying states – when interacting with its environment. Deploying distinct QRFs breaks the thermodynamic symmetry of the screen for each agent, redirecting energy flows within each agent to fund processing and recording memory records of some bits while others – those in the "thermodynamic" sector $F$ – remain necessarily unobserved. Thus the quantum formalism explicitly enforces a distinction that the classical formalism leaves implicit: that between observed and unobserved parts of the environment. Mismatches between QRFs deployed by two interacting agents generate noise and prediction errors, including incorrect Frame Problem solutions and failures to correctly re-identify objects.

In the classical setting, agents "self-evidence" by maintaining their nonequilibrium steady-states or, equivalently, the integrity of their MBs. From a mathematical point of view, this is maintaining their identities (in the sense of being independently well-defined) as systems. In the quantum setting, being independently well-defined is being *separable* from – not entangled with – the environment; hence "self-evidencing" is maintaining separability. The FEP identifies self-evidencing with the minimization of Bayesian prediction error: to "be" is to be capable of successful predictions; sometimes described as "predicting yourself into existence". Minimizing prediction error is, in the quantum setting, minimizing the difference between two Markov kernels. As in the classical setting, noise and errors due to insufficient learning must both be minimized, a process that requires trade-offs between the two.

The classical and quantum formulations of the FEP differ in their asymptotic behavior, i.e., as total prediction error is driven toward zero. A classical "perfect predictor" achieves maximal classical correlation and hence maximal behavioral synchrony with its environment; it becomes a "perfect regulator" in the sense of the good regulator theorem [148]. In a quantum setting, perfect prediction entails shared QRFs between interaction partners and hence entanglement; a quantum "perfect regulator" becomes indistinguishable from the environment it is predicting. While this difference in asymptotic behaviors is a formal, mathematical outcome, it can be traced to a difference in fundamental assumptions. The classical formalism assumes a spacetime background, and hence can rely on separation in space to distinguish between systems. The quantum formalism is background free: space is simply an observable, represented by a QRF that a system may or may not deploy. It can, therefore, play no "ontic" role in maintaining distinctions between systems. This re-

flects the general role of ("physical" 3d) space in quantum field theories: space is there to enforce separability (see [77] for a general discussion of this point from a gauge-theoretic perspective).

Both classical and quantum formulations of the FEP engender a fundamental form of *epistemic* solipsism, in the sense that the coupling between internal and external systems precludes the states of one from ever "knowing" the states of the other. While this seems counter-intuitive, it is a straightforward consequence of the use of vector spaces to represent physical states. Vector-space product operators – the classical Cartesian product or the Hilbert-space tensor product – are by definition associative (equivalently, dynamic operators such as Hamiltonians are additive); product decompositions are, therefore, undetectable across decompositional boundaries in any vector space [149]. This sense of epistemic solipsism does *not*, as Fuchs emphasizes [64], in any way suggest ontic or metaphysical solipsism. Both classical and quantum formulations of the FEP – indeed, any theory of measurement or observation – requires two interacting systems, observer and observed. The idea of a metaphysically solipsist theory of observation is self-contradictory.

## 5.2   Summary of results

We have obtained three results in this paper:

1. Given the standard free-choice assumption, the intuitive idea of an "agent" or IGUS can be fully formulated within background-independent, scale-free quantum information theory.

2. The FEP can be given a quantum-theoretic formulation that renders it applicable to generic quantum systems.

3. When formulated as a generic principle of quantum information theory, the FEP is asymptotically equivalent to the Principle of Unitarity.

Result 1) places the long-standing practice of treating generic quantum systems as "observers" on a firm theoretical foundation. It allows a formal specification, in the language of QRFs, of exactly what systems a given observer is capable of recognizing and what states of those systems it is capable of measuring. Such specifications require no "ontic" assumptions of observer-independent systems; hence they are compliant with a device-independent theoretical strategy. Result 1) also provides a formal definition of an "agent" in the context of the free-choice assumption, and provides a generic language – the category-theoretic language of Channel Theory – for specifying the semantics assigned by an agent to an observational outcome. Finally, Result 1) shows how semantics arise from thermodynamic symmetry breaking on a holographic screen, and provides a formal mechanism for quantifying energy flows that enable classical computation and classical memory encoding. As the underlying quantum information theory is scale-free, this symmetry-breaking mechanism

is scale-free; hence it can be employed by biological systems at any scale, from molecular-interaction networks to ecosystems.

Result 2) extends the range of applicability of the FEP to generic quantum systems independently of spacetime background or scale-dependent assumptions. Quantum fields, black holes, and other spatially-distributed or topologically-defined systems can, therefore, be regarded as Bayesian observers and, if free choice is assumed, as Bayesian agents. Result 2) therefore makes clear the sense in which generic quantum systems can be regarded as "users" of quantum information theory, as proposed under the rubric of QBism [42, 44, 64]. Indeed Result 2) renders QBism a consequence of quantum information theory, not an interpretation.

Result 3) shows that the FEP is compliant with the Principle of Unitarity and, conversely, that unitary evolution is compliant with the FEP. It allows us, in particular, to understand quantum context-switching as both a source of prediction errors and a strategy for reducing prediction errors. Result 3) also allows us to view separability as a resource for classical communication and computation that is analogous to entanglement as a resource for quantum communication and computation. Hence, it allows us to consider trade-offs between these resources by systems that maintain approximate separability while also employing shared entanglement. Such systems have been studied in the abstract, and can potentially exceed the computational power of Turing machines (e.g. can solve the Halting problem as shown in [150]). Result 3) suggests that (some) biological systems may have this capability, as discussed further in §5.4 below.

## 5.3 Applications to biological cognition

In addition to the results listed above, the current framework has a variety of more specifically biological consequences, some of which have been discussed already in [13]. It predicts, for example, that moving in ordinary 3d space does not require a QRF for Euclidean space and hence, does not require an experience of space. From a classical perspective, this is certainly true; as evidenced e.g. by place and grid cells in mammalian brains [151, 152, 153] that appear to encode a coarse-grained representation of location, head-direction etc., in various frames of reference. This coarse graining endorses another prediction; namely, that actionable classical encodings are coarse-grained. Any system that encodes information irreversibly is, therefore, faced with a choice that its computational architecture must resolve: the trade-off between preserving information in memory and losing information due to coarse graining. On a classical FEP account, this coarse graining is mandated by the minimization of VFE, which can be expressed as complexity minus accuracy. Classically, complexity corresponds to the KL divergence between posterior and prior beliefs as in Eq. (35). This relative entropy clearly depends upon the degree of discretisation or coarse graining afforded by the dimensionality of internal states. Effectively, this KL divergence scores the computational and thermodynamic cost of belief updating that is mitigated by coarse graining or, in quantum terms, devoting memory resources to the results computed by only some QRFs, possibly in a context-sensitive fashion [154].

43

Modeling QRFs with CCCDs suggests that the intrinsically hierarchical nature of CCCDs will be generically inherited by QRFs. Hence the current framework predicts that living systems will generically organize information processing hierarchically. This is observed in living systems at all scales, and applies to both direct interactions with the environment and the operation of memory systems. Biological memory systems, for example, are hierarchical from the macromolecular scale upwards, spanning at least ten orders of magnitude in spatial scale and upwards of 15 orders of magnitude in temporal scale [155]. The implementation of sensation and action QRFs by unicellular organisms is discussed in [13, 86]; the implementation of hierarchies of QRFs by cortical neurons, from the scales of individual dendritic spines through dendritic branches to whole neurons, is discussed in [156]. This latter model merges naturally with models of QRFs implemented by functional neural networks, up to the scale of global-workspace models [101, 102]. Hierarchy induction can be expected whenever coarse-graining, and hence abstraction of a general actionable property from an equivalence class of instances, can be expected to increase information-processing efficiency, particularly its free-energy efficiency. Selective constraints operating at multiple time scales provably induce regulatory (i.e. metaprocessor) hierarchies under generic conditions, provided only that the overhead cost of maintaining the hierarchy is sufficiently small [157]. The practical implications of functional (information-processing) hierarchy, in addition to the obvious structural hierarchy, of levels in biology are discussed for example in [158, 159].

The ubiquity of context-dependent effects leads to another prediction: living systems in complex environments will evolve context and attention switching systems. On a classical view, this entails the identification of context (cf. the role of pointers) in a hierarchical generative model, where high level states contextualise the processing of lower level states. This immediately introduces the notion of MBs or holographic screens in the joint space of an agent's internal states, i.e. a notion of modularity supported by shared memory. This is gracefully accommodated by the channel theory of QRFs as developed in §2.6.[12] Indeed, much work in the classical field of active inference rests upon optimising the hierarchical structure of deep generative models, via various free energy minimizing processes [161, 162, 163, 164]. This is especially true for generative models of navigation and language [165, 166, 167] that are almost universally based upon quantisation or discretisation of state spaces. The attendant Boolean logic that arises from the imposition of Boolean constraints by QRFs may have important practical applications, as demonstrated by recent work in active vision and scene construction [168] that rests upon a generative model of the sensory (visual) consequences of visually foraging a scene with multiple objects. Active inference in this context can be seen as an inversion of a generative model that maps from causes (external objects), to consequences (sensory states) (cf. [129]).

The inversion of generative models for active vision is extremely difficult and ill-posed due to its computational complexity. These models have to accommodate all the natural physical laws of motion and optics (e.g., occlusion). The inversion of these models corresponds to

---

[12]See [102] for an explicit application to global-workspace theory and [160] for a more ambitious synthesis of active inference, a global workspace, and IIT.

the measurement operators that mediate belief updating. In a classical setting, this can be massively finessed by coarse graining the problem and applying Boolean operators. For example, if one object is near and another object is far, an agent will see the near object. Quantising or coarse graining the internal (i.e. generative) model along these lines reduces the likelihood mapping to a set of relatively simple Boolean operators that could be cast as measurement operators in a quantum-theoretic context. At present, when these operators are encoded on standard classical computers for simulation, they are effectively implemented with enormous tensors. Reading and writing these tensors into working memory dwarfs the actual compute time and renders active vision schemes of this sort computationally and thermodynamically inefficient. One might imagine that a combination of quantum computing [34, 36] and neuromorphic engineering [169, 170] may be able to parallel the efficiency of human vision.

There are, in addition, certain technical problems that are resolved in moving from a classical to a quantum FEP formulation of active inference that go beyond a commitment to unitary processes and binary measurement operators. These include problems entailed by assuming reference frames that can be specified to infinite precision. In formalising the asymptotic behaviour of belief updating in the absence of random fluctuations in the classical formalism, for example, one has to deal with differential entropies that are ill-defined for very precise conditional densities, e.g. Dirac Delta functions. One workaround is to use Jaynes' limiting density of discrete points (LDDP) [171], which brings us back to where we started, namely finite dimensional Hilbert spaces and discrete state spaces.

Finally, from a more philosophical perspective, the framework presented here is consistent with [13, 14, 172] in supporting a panpsychist perspective on questions of agency, sentience, and cognition. As our Definition 1 of agency illustrates, traditional binary categorizations of entities into those having agency, sentience, and cognition and those lacking them are replaced here by continua of "interestingness" along these dimensions. Definition 1 assumes free choice, in particular of measurement basis. The Conway-Kochen theorem [81] states that if *any* physical system is assumed to have free choice, then *all* physical systems must be assumed to have free choice. "Free choice" in this case means behavior that is not determined by (cannot be fully predicted given) the events in the system's past lightcone. Determination of behavior by events in the past lightcone is *local* determinism; such local determinism is fundamentally inconsistent with the global determinism implied by unitarity [83]. Hence what is interesting is the *extent* of choice. As emphasized in [10], interesting choices are implemented by internal processes. An electron, for example, has internal states – its charge and mass are not states of its MB – but they are invariant, and so do not implement choices. The internal states of rocks are not invariants – changes in water content or radioactive decay can occur – but they do not, in general, implement interesting choices from our human perspective. Interesting choices require interesting sensations and actions, and hence significant internal energy flows as discussed above. They require differential *use* of thermodynamic resources to deploy multiple QRFs that probe the observable environment in different ways. Systems, including organisms, clearly do this to different extents; even among humans, the extent of variation is striking. Consonant with a broadly-construed

enactivism [173, 174, 175, 176], we view such active exploration of the environment – hence, active inference – as indicating cognition and intelligence.

## 5.4 Predictions and open questions

The results obtained here suggest that the field of quantum biology is far larger than has so far been explored; indeed they suggest that all biological systems are properly considered quantum systems and can be expected to employ quantum coherence as an information processing resource. This suggestion is of course not novel, having been explored by many authors from philosophical, theoretical, and increasingly over the past two decades, empirical perspectives [177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191]. We have, however, shown here that it follows from general principles: quantum systems with sufficiently complex internal information flows can be expected to exhibit active inference, and hence to behave like organisms. Indeed, quantum systems that do *not* display evident intelligence – quantum systems that are trivial agents – become an unusual special case. Conversely, the results obtained here suggest that the concepts of active inference, agency, Bayesian satisficing, and cognition are applicable to generic quantum systems, and hence to physical systems across the board. This turns the traditional question of the emergence of intentionality [192, 193, 194] on its head: by making agency a generic expectation, it makes the "merely physical" the special case demanding explanation. It thus suggests that the traditional division between agency and mechanism is a hindrance, not a help, in the task of understanding the natural world.

We make, in particular, the following predictions:

1. The internal, molecular-scale dynamics of both prokaryotic and eukaryotic cells implement quantum information processing, i.e., make essential use of quantum coherence as a computational and memory resource. This is consistent with recent results showing that the free energy budgets of biological systems across known phylogeny are orders of magnitude short of the resources required for purely classical computation at the molecular scale [195].

2. Interacting biological systems trade off separability against entanglement and hence classical against quantum communication. Interactions between biological systems from the molecular scale upwards can be expected to display quantum contextuality and violations of the Bell and Leggett-Garg inequalities.

These predictions remain largely untested, both for reasons of technical difficulty and due to a still-pervasive traditional view of macroscopic systems as ontically classical. The mechanistic role of entanglement – even in relatively well-established systems such as light harvesting – remains subject to considerable debate [189, 190, 191]. Quantum context switching has been detected in human subjects [196, 197], but whether it is properly considered "quantum"

or merely "quantum-like" remains open to question [198].[13] Neither Bell nor Leggett-Garg inequality violations have been conclusively demonstrated with biological systems.

Setting technical difficulties aside, we suggest that coherence effects in biological settings may be systematically overlooked due to being dismissed as "noise" or "random coincidence." Thermal noise in biological settings is well characterized; general environmental variation, including effects of signaling by other cells or organisms, is less straightforwardly modeled or controlled. Experimental designs that explicitly test for coherence effects will, we expect, be required to test the above predictions. Focused theoretical support for such designs is therefore necessary.

Interactions involving multiple agents remain an open theoretical as well as experimental-design problem. In the bipartite setting employed for the technical analysis above, every agent interacts with its entire surrounding, whether the bits transferred by this interaction are observed and processed or not (i.e. whether they are included in sector $E$ or $F$). This does not change in a multi-party setting; each agent still interacts with its entire surround, identifying other agents (or not) via specific QRFs. The ubiquitous assumption that inter-agent communication is classical, made in domains as disparate as cell-cell interaction and human natural language use, becomes problematic in this setting. As discussed in §4.3, shared QRFs are required for fully-shared, counterfactual-supporting semantics, but they induce entanglement (see also the discussion of this point in [47]). The extent of shared semantics is not readily observable in living systems. In a fully classical setting, generalised synchronisation (a.k.a., synchronisation of chaos) emerges when two free energy minimising 'partners' observe each other [200]. "Perfect prediction" of the partner's behavior may result, driven by classical learning of a shared generative model, implicitly resolving the hermeneutics problem in the communication context [201, 202]. Recognizing a particular communication partner, in this case, requires invoking a particular generative model, the correct model to predict that partner's behavior [203]. Such classical synchronization is not, however, robust against perturbation; altering one agent's model does not "automatically" alter the other, as would be expected if the models, i.e. the QRFs implemented by the two agents were entangled. Both theoretical and experimental characterization of relevant QRFs will be required to assess the extent to which classical communication can be considered purely classical, and determine where and how quantum coherence contributes to in-practice successful shared semantics.

Darwinian evolution can be viewed as a process of variation and selection of QRFs, and hence as an instance of the multiple-agents problem in which semantics is only partially shared. Evolution can be given a natural description in the framework of the classical FEP [5, 204, 205, 206]. Variation and selection have been advanced as a model of decoherence under the rubric of quantum Darwinism [27, 207]; see [208] for a discussion from the perspective of universal Darwinism. The selection mechanism invoked by quantum Darwinism, however, assumes QRF sharing by multiple agents [209]; see [47, 71] for discussion. While

---

[13]Such "quantum-like" contextuality is claimed for a certain case of gene expression in [199] where conflicting probabilities (quantum vs. classical) do not give rise to a consistently definable joint distribution, so suggesting a variant of Kolmogorov contextuality.

the present results allow any evolutionary system coupled to a larger environment to be viewed as a Bayesian agent implementing active inference, a fully-satisfactory account of variation and selection within a quantum framework remains to be developed.

Additional theoretical work is also needed to understand the relationships among the many distinct models of quantum contextuality that have been put forward, often with the use of quite different formal tools (e.g. [210, 211, 212, 213]). The question of "context" is deeply tied up with that of what is to be regarded as the "environment" or "surrounding" of any given system. This is a particularly critical question at the cellular level, where multiple signaling modalities with different spatial ranges and temporal characteristics are present. Models that explicitly characterize the "spaces" in which cells and multicellular systems operate – e.g. the space of potential morphologies, or that of potential messages from interaction partners – and that consider constraining effects of one space on another both within and between scales will be needed to understand the roles of context, and of context switching, in biological systems.

In closing, we hope that we have shown to readers familiar with the FEP and the active inference framework that quantum effects are worth considering both theoretically and in experimental design. For readers not familiar with the classical FEP but literate in quantum theory (or vice versa), we hope this paper has gone some way to contextualizing your QRFs in sense-making via Markov blankets and their underlying holographic screens.

# Conflict of Interest Statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

# Funding

# Acknowledgement

# References

[1] Friston KJ. A theory of cortical responses. Philos Trans R Soc Lond B, Biol Sci 2005;360:815–36.

[2] Friston KJ, Kilner J, Harrison L. A free energy principle for the brain. J Physiol Paris 2006;100:70–87.

[3] Friston KJ, Stephan KE. Free-energy and the brain. Synthese 2007;159:417–58.

[4] Friston KJ. The free-energy principle: a unified brain theory? Nat Rev Neurosci 2010;11:127–38.

[5] Friston KJ. Life as we know it. J R Soc Interface 2013;10:20130475.

[6] Friston KJ, FitzGerald T, Rigoli F, Schwartenbeck P, Pezzulo G. Active inference: a process theory. Neural Comput 2017;29:1–49.

[7] Ramstead MJ, Badcock PB, Friston KJ. Answering Schrödinger's question: a free-energy formulation. Phys Life Rev 2018;24:1–16.

[8] Ramstead MJ, Constant A, Badcock PB, Friston KJ. Variational ecology and the physics of sentient systems. Phys Life Rev 2019;31:188–205.

[9] Kuchling F, Friston K, Georgiev G, Levin M. Morphogenesis as Bayesian inference: A variational approach to pattern formation and control in complex biological systems. Phys Life Rev 2020;33:88–108

[10] Friston K. A free-energy principle for a particular physics. Preprint arxiv:1906.10184 (2019).

[11] Clark A. How to knit your own Markov blanket: Resisting the Second Law with meta-morphic minds. In: Metzinger T, Wiese W (eds), Philosophy and Predictive Processing: 3. Frankfurt am Main: MIND Group, 2017.

[12] Scholl BJ, Tremoulet PD. Perceptual causality and animacy. Trends Cogn Sci 2000;4:299–309.

[13] Fields C, Glazebrook JF, Levin M. Minimal physicalism as a scale-free substrate for cognition and consciousness. Neurosci Cons 2021;7:niab013.

[14] Levin M. Life, death, and self: fundamental questions of primitive cognition viewed through the lens of body plasticity and synthetic organisms. Biochem Biophys Res Comm 2020;564:114–133.

[15] Pylyshyn, Z. Computation and Cognition: Toward a Foundation for Cognitive Science. Cambridge, MA: MIT Press, 1984.

[16] Church, A. A Note on the *Entscheidungsproblem.* J. Symbol. Logic 1936;1:40–41.

[17] Turing, A. M. On computable numbers, with an application to the *Entscheidungsproblem.* Proc. London Math. Soc. Ser. 2 1936;42:230–265.

[18] Putnam, H. Psychological predicates. In Capitan, W. and Merril, D. (Eds) Art, Mind, and Religion. Pittsburgh, PA: University of Pittsburgh Press, pp. 37–48.

[19] Fodor, J. Psychological Explanation. Cambridge, MA: MIT Press, 1968.

[20] Ashby, W.R. Introduction to Cybernetics. London: Chapman and Hall, 1956.

[21] Moore, E. F. Gedanken experiments on sequential machines. In: Shannon, C. W.; McCarthy, J. (Eds.) Autonoma Studies. Princeton, NJ: Princeton University Press, pp. 129–155.

[22] Booch G, Maksimchuk RA, Engle MW, Young BJ, Conallen J, Houston KA. Object-Oriented Analysis and Design with Applications, 3rd Ed. Upper Saddle River, NJ: Addision-Wesley, 2007

[23] Laughlin RB, Pines D. The theory of everything. Proc. Natl. Acad. Sci. USA 2000;97:28–31.

[24] Von Neumann J. The Mathematical Foundations of Quantum Mechanics. Princeton, NJ: Princeton University Press, 1955.

[25] Omnès R. Consistent interpretations of quantum mechanics. Rev Mod Phys 1992;64:339–382.

[26] Zurek WH. Decoherence, einselection and the existential interpretation (the rough guide). Phil Trans R Soc A 1998;356:1793–1821.

[27] Zurek WH. Decoherence, einselection, and the quantum origins of the classical. Rev Mod Phys 2003;75;715–775.

[28] Schlosshauer M. Decoherence, the measurement problem, and interpretations of quantum mechanics. Rev Mod Phys 2003;76:1267–1305.

[29] Schlosshauer M. Decoherece and the Quantum to Classical Transition. Springer, Berlin, 2007.

[30] Bassi A, Lochan K, Satin S, Singh TJ, Ulbricht H. Models of wave-function collapse, underlying theories, and experimental tests. Rev Mod Phys 2013;85:471–527.

[31] Landsman NP. Between classical and quantum. In: Butterfield J, Earman J. (eds.), HandBook of the Philosophy of Science: Philosophy of Physics. Elsevier, Amsterdam, Netherlands, 2007, pp. 417–553.

[32] Cabello A. Interpretations of quantum theory: A map of madness. Preprint arxiv:1509.04711v1, 2015.

[33] Wheeler JA. Law without law. In: Wheeler JA, Zurek W (eds), Quantum Theory and Measurement. Princeton, NJ: Princeton University Press, 1983, pp. 182–213.

[34] Feynman RP. Simulating physics with computers. Int J Theor Phys 1982;21:467–488.

[35] Deutsch D. Quantum theory, the Church-Turing principle and the universal quantum computer. Proc R Soc A 1985;400: 97–117.

[36] Nielsen MA, Chuang IL. Quantum Computation and Quantum Information. New York: Cambridge University Press, 2000.

[37] Hardy L. Quantum theory from five reasonable axioms. Preprint arxiv:quant-ph/0101012v4 (2001).

[38] Fuchs CA. Quantum mechanics as quantum information, mostly. J Mod Opt 2003;50:987–1023.

[39] Brassard G. Is information the key? Nat Phys 2005;1:2–4.

[40] Chiribella G, D'Ariano GM, Perinotti P. Informational derivation of quantum theory. Phys Rev A 2011;84:012311.

[41] Masanes L, Müller MP. A derivation of quantum theory from physical requirements. New J Phys 2011:13;063001.

[42] Fuchs C, Schack R. Quantum Bayesian coherence. Rev Mod Phys 2013;85:1693–1715.

[43] Grinbaum A. How device-independent approaches change the meaning of physical theory. Stud Hist Phil Mod Phys 2017;58:22–30.

[44] Mermin ND. Making better sense of quantum mechanics. Rep Prog Phys 2018;82:012002.

[45] Müller MP. Law without law: from observer states to physics via algorithmic information theory. Quantum 2020;4:301.

[46] Fields C, Glazebrook JF. Representing measurement as a thermodynamic symmetry breaking. Symmetry 2020;12:810.

[47] Fields C, Glazebrook JF, Marcianò A. Reference frame induced symmetry breaking on holographic screens. Symmetry 2021;13:408.

[48] Aharonov Y, Kaufherr T. Quantum frames of reference. Phys Rev D 1984;30:368–385.

[49] Bartlett SD, Rudolph T, Spekkens RW. Reference frames, super-selection rules, and quantum information. Rev Mod Phys 2007;79:555–609.

51

[50] Adámek J, Herrlich H, Strecker GE. Abstract and Concrete Categories: The Joy of Cats. New York: Wiley, 2004. Available at http://katmat.math.uni-bremen.de/acc (Accessed May 26, 2019)

[51] Awodey S. Category Theory. (Oxford Logic Guides, 62). Oxford, UK: Oxford University Press, 2010.

[52] Barwise J, Seligman J, Information Flow: The Logic of Distributed Systems (Cambridge Tracts in Theoretical Computer Science, 44). Cambridge, UK: Cambridge University Press, 1997.

[53] Kochen S, Specker EP. The problem of hidden variables in quantum mechanics. J Math Mech 1967;17:59–87.

[54] Mermin D. Hidden variables and the two theorems of John Bell. Rev Mod Phys 1993;65:803–815.

[55] Gell-Mann M, Hartle JB. Quantum mechanics in the light of quantum cosmology. In Zurek W (ed) Complexity, Entropy, and the Physics of Information. Boca Raton, FL: CRC Press, 1989; pp. 425–458.

[56] Bell JS. On the Einstein-Podolsky-Rosen paradox. Physics 1964;1:195–200

[57] Emary C, Lambert N, Nori F. Leggett-Garg inequalities. Rep Prog Phys 2013;77:016001.

[58] Landauer R. Irreversibility and heat generation in the computing process. IBM J Res Devel 1961;5:183–195.

[59] Landauer R. Information is a physical entity. Physica A 1999;263:63–67.

[60] Bennett CH. The thermodynamics of computation. Int J Theor Phys 1982;21:905–940.

[61] Lloyd S. Ultimate physical limits to computation. Nature 2000;406:1047–1054.

[62] Zweir MC, Chong LT. Reaching biological timescales with all-atom molecular dynamics simulations. Curr Opin Pharmacol 2010;10:745–752.

[63] Wang Q, Schoenlein RW, Peteanu LA, Mathies RA, Shank CV. Vibrationally coherent photochemistry in the femtosecond primary event of vision. Science 1994;266:422–424.

[64] Fuchs C. QBism, the Perimeter of quantum Bayesianism. Preprint arxiv:1003.5209, 2010.

[65] Moretti V, Oppio M. Quantum theory in real Hilbert space: How the complex Hilbert space structure emerges from Poincaré symmetry. Rev Math Phys 2017;29:17500021.

[66] Renou M-O, Trillo D, Weilenmann M, Thinh LP, Tavakoli A, Gisin N, Acín A, Navascués M. Quantum physics needs complex numbers. Preprint arxiv:2101.10873, 2021.

[67] Baez JC. Getting to the bottom of Noether's theorem. Preprint arxiv:2006.14741, 2020.

[68] Kauffman LH. Eigenforms and quantum physics. Cybernet Human Knowing 2011;18:111–121.

[69] Aspect A, Grangier P, Roger G. Experimental realization of the Einstein-Posolsky-Rosen *gedankenexperiment*: A new violation of Bell's inequalities. Phys Rev Lett 1982;49:91–94.

[70] Fields C, Marcianò A. Sharing nonfungible information requires shared nonfungible information. Quant Rep 2019;1: 252–259.

[71] Fields C, Marcianò A. Holographic screens are classical information channels. Quant Rep 2019;2;326–336.

[72] Wheeler JA. Information, physics, quantum: The search for links. In Zurek W (ed) Complexity, Entropy, and the Physics of Information. Boca Raton, FL: CRC Press, 1989; pp. 3–28.

[73] 't Hooft G. Dimensional reduction in quantum gravity. In Ali A, Ellis J, Randjbar-Daemi S. (eds) Salamfestschrift. Singapore: World Scientific, 1993, pp. 284–296.

[74] Susskind L. The world as a hologram. J Math Phys 1995;36:6377–6396.

[75] Bousso R. The holographic principle. Rev Mod Phys 2002;74:825–874.

[76] Almheiri A, Hartman T, Maldacena J, Shaghoulian E, Tajdini A. The entropy of Hawking radiation. Rev Mod Phys 2021;93:035002.

[77] Addazi A, Chen P, Fabrocini F, Fields C, Greco E, Lulli M, Marcianò A, Pasechnik R. Generalized holographic principle, gauge invariance and the emergence of gravity à la Wilczek. Front Astron Space Sci 2021;8:563450.

[78] Fields C, Marcianò A. Markov blankets are general physical interaction surfaces. Phys Life Rev 2020;33:109–111.

[79] Bohr N. Atomic Physics and Human Knowledge. New York: Wiley, 1958.

[80] Gisin N. Non-realism: Deep thought or a soft option? Found Phys 2012;42:80–85.

[81] Conway JH, Kochen S. The strong free will theorem. Notices AMS 2009;56(2):226–232.

[82] Tipler, F. Quantum nonlocality does not exist. Proc. Natl. Acad. Sci. USA 2014;111:11281–11286.

[83] Fields C. A whole box of Pandoras: Systems, boundaries and free will in quantum theory. J Expt Theor Artif Intell 2013;25:291–302.

[84] Horsman C, Stepney S, Wagner RC, Kendon V. When does a physical system compute? Proc R Soc A 2014;470:20140182.

[85] Rice HG. Classes of recursively enumerable sets and their decision problems. Trans Am Math Soc 1953;74:358–366.

[86] Fields C, Levin M. How do living systems create meaning? Philosophies 2020;5:36.

[87] Quive WVO. Word and Object. Cambridge, MA: MIT Press, 1960.

[88] Zanardi, P. Virtual quantum subsystems. Phys Rev Lett 2001;87:077901.

[89] Zanardi P, Lidar DA, Lloyd S. Quantum tensor product structures are observable-induced. Phys Rev Lett 2004;92:060402.

[90] Dugić M, Jeknić J. What is "system": Some decoherence-theory arguments. Int J Theor Phys 2006;45:2249–2259.

[91] Dugić M, Jeknić-Dugić J. What is "system": The information-theoretic arguments. Int J Theor Phys 2008;47:805–813.

[92] De la Torre AC, Goyeneche D, Leitao L. Entanglement for all quantum states. Eur J Phys 2010;31:325–332.

[93] Harshman NL, Ranade KS. Observables can be tailored to change the entanglement of any pure state. Phys Rev A 2011;84:012303.

[94] Buhrmann T, Di Paolo E. The sense of agency - a phenomenological consequence of enacting sensorimotor schemes. Phenom Cog Sci 2017;16:207–236

[95] Barr M. *-Autonomous Categories, with an Appendix by Po Hsiang Chu (Lecture Notes in Mathematics 752). Springer: Berlin, Germany, 1979.

[96] Pratt V. Chu spaces. In School on Category Theory and Applications (Coimbra 1999); Volume 21 of Textos Math. Sér. B. University of Coimbra: Coimbra, Portugal, 1999, pp. 39–100.

[97] Pratt V. Chu spaces from the representational viewpoint. Ann Pure Appl Logic 1999;96:319–333.

[98] Fields C, Glazebrook JF. A mosaic of Chu spaces and Channel Theory I: Category-theoretic concepts and tools. J Expt Theor Artif Intell 2019;31:177–213.

[99] Fields C, Glazebrook JF. Information flow in context-dependent hierarchical Bayesian inference. J Expt Theor Artif Intell 2022;34:111–142.

[100] Collier J. Information, causation and computation. In G. D. Crnkovic and M. Burgin (eds.) Information and Computation: Essays on Scientific and Philosophical Foundations of Information and Computation (World Scientific Series in Information Studies Vol 2). World Scientific Press. Hackensack, NJ, 2011, pp. 89-105.

[101] Fields C, Glazebrook JF. A mosaic of Chu spaces and Channel Theory II: Applications to object identification and mereological complexity. J Expt Theor Artif Intell 2019;31:237–265.

[102] Fields C, Glazebrook JF. Do Process-1 simulations generate the epistemic feelings that drive Process-2 decision making? Cogn Proc 2020;21:533–553.

[103] Barwise J. Information and impossibilities. Notre Dame J Formal Logic 1997;38(4):488–515.

[104] Cherniak E. Bayesian networks without tears. AI Mag 1991;12(4):50–63.

[105] Allwein, G. A qualititative framework for Shannon Information theories. In NSPW '04: Proceedings of 2004 Workshop on New Security Paradigms (Nova Scotia, Canada, September 20-23, 2004). New York: ACM, 2004, pp. 23–31.

[106] St. Clere Smithe T. Compositional active inference I: Bayesian lenses, statistical games. Preprint arxiv:2109.04461 [math.ST], 2021.

[107] Coecke B. Quantum picturalism. Contemp Phys 2010;51:59–83.

[108] Adams EW. A Primer of Probabilistic Logic. Chicago: University of Chicago Press, 1998.

[109] Dzhafarov EN, Cervantes VH, Kujala JV. Contextuality in canonical systems of random variables. Phil Trans R Soc A 2017;375:20160389.

[110] Dzharfarov EN, Kon M. On universality of classical probability with contextually labeled random variables. J Math Psych 2018;85:17–24.

[111] List C. Why Free Will is Real. Harvard Univ. Press, Cambridge MA, 2019.

[112] Ellis GFR. Top-down causation and emergence: some comments on mechanisms. In, Interface Focus: a theme supplement of Journal of the Royal Society Interface 2011;2(1): 126–140.

[113] Ellis GFR. Emergence in solid state physics and biology. Foundations in Physics 2020;50: 1098–1139

[114] MacKay DJC, Peto LCB. A hierarchical Dirichlet language model. Nat Lang Engin 1995;1:289–308.

[115] Wallace CS, Dowe, DL. Minimum message length and Kolmogorov complexity. Computer J 1999;42:270–283.

[116] Smith R, Schwartenbeck P, Parr T, Friston KJ. An active inference approach to modeling structure learning: Concept learning as an example case. Front Comput Neurosci 2020;14:41.

[117] Weinstein A. Groupoids: Unifying internal and external symmetry. Notices Am Math Soc 1996;43:744–752.

[118] Brown R. *Topology and Groupoids*. www.groupoids.org.uk, Deganwy, UK, 2006.

[119] Deutsch D. The structure of the multiverse. Proc R Soc A 2002;458:2911–2923.

[120] Marcianò, A.; Chen, D.; Fabrocini, F.; Fields, C.; Greco, E.; Gresnigt, N.; Jinklub, K.; Lulli, M., Terzidis, K.; Zappala, E. Deep neural networks as the semi-classical limit of quantum neural networks. Preprint arxiv:2007.00142v2 [cond-mat.diss-nn].

[121] Oizumi M, Albantakis L, Tononi G. From the phenomenology to the mechanisms of consciousness: Integrated Information Theory 3.0. PLoS Comp Biol 2014;10:e1003588.

[122] Robbins RJ, Krishtalka L, Wooley JC. Advances in biodiversity: Metagenomics and the unveiling of biological dark matter. Stand Genom Sci 2016:11:69.

[123] Scholl BJ. Object persistence in philosophy and psychology. Mind Lang 2007;22:563–591.

[124] Fields C. The very same thing: Extending the object token concept to incorporate causal constraints on individual identity. Adv Cognit Psychol 2012;8:234–247.

[125] Fields C. Some consequences of the thermodynamic cost of system identification. Entropy 2018;20:797.

[126] Pais A. Einstein and the quantum theory. Rev Mod Phys 1979;51:863–914.

[127] Marr D. Vision. Freeman, San Francisco, CA, 1982.

[128] Palmer S. Vision Science: Photons to Phenomenology. MIT Press, Cambridge, MA, 1999.

[129] Pizlo Z. Perception viewed as an inverse problem. Vis Res 2001;41:3145–3161.

[130] Prakash C, Fields C, Hoffman DD, Prentner R, Singh M. Fact, fiction, and fitness. Entropy 2020;22:514.

[131] Prakash C, Stephens KD, Hoffman DD, Prentner R, Singh M, Fields C. Fitness beats truth in the evolution of perception. Acta Biotheor 2021;69: 319–341.

[132] Bohr N. The quantum postulate and the recent development of atomic theory. Nature 1928;121:580–590.

[133] Di Biagio A, Donà P, Rovelli C. The arrow of time in operational formulations of quantum theory. Quantum 2021;5;520.

[134] Tegmark M. How unitary cosmology generalizes thermodynamics and solves the inflationary entropy problem. Phys Rev D 2012;85:123517.

[135] Sulis W. Contextuality in neurobehavioural and collective intelligence systems. Quantum Reports 2021;3: 592–614.

[136] Mermin D. Simplified unified form for the major no-hidden variables theorem. Phys Rev Lett 1990; 65: 3373–3376.

[137] McCarthy J, Hayes PJ. Some philosophical problems from the standpoint of artificial intelligence. In Michie D, Meltzer, B. (eds.) Machine intelligence, Vol. 4). Edinburgh University Press, Edinburgh, 1969, pp. 463–502.

[138] Fields C. How humans solve the frame problem. J Expt Theor Artif Intell 2013;25:441–456.

[139] Gottlieb J, Lopes M, Oudeyer P-Y. Motivated cognition: Neural and computational mechanisms of curiosity, attention, and intrinsic motivation. In: Kim S-Y, Reeve J, Bong M. (Eds) Recent Developments in Neuroscience Research on Human Motivation (Advances in Motivation and Achievement, Vol. 19), Emerald Group Publishing Limited, Bingley, UK, 2016, pp. 149–172.

[140] Wooters WK, Zurek WH. A single quantum cannot be cloned. Nature 1982;299:802–803.

[141] Pratt V. Types as Processes, via Chu spaces. Elect Notes Theor Comp Sci 1997;7:227–247.

[142] Pratt V. Broadening the denotational semantics of linear logic. Elect Notes Theor Comp Sci 1996;3: 155–166.

[143] Culbertson J, Sturtz K. Bayesian machine learning via category theory. Preprint arXiv:1312.1445v1[math.CT], 2013.

[144] Culbertson J, Sturtz K. A categorical foundation for Bayesian probability. Appl Categor Struct 2014;22(4): 647–662.

[145] Giry M. A categorical approach to probability theory. In: Banaschewski B. (ed) Categorical Aspects of Topology and Analysis. (Lecture Notes in Mathematics, vol 915). Springer, Berlin, 1982, pp. 68–85.

[146] Lawvere WF. The category of probabilistic mappings. Unpublished seminar notes, (1962) available at https://ncatlab.org/nlab/files/lawvereprobability196.pdf (Accessed 10 Dec 2021).

[147] Abramsky S. Big toy models: Representing physical systems as Chu spaces. Synthese 2012;186:697–718.

[148] Conant RC, Ashby WR. Every good regulator of a system must be a model of that system. Int J Syst Sci 1970;1(2):89–97.

[149] Fields C. Building the observer into the system: Toward a realistic description of human interaction with the world. Systems 2016;4:32.

[150] Ji Z, Natarajan A, Vidick T, Wright J, Yuen H. MIP* = RE. Comms ACM 2021;64(11):131–138.

[151] Moser MB., Rowland DC, Moser EI. Place cells, grid cells, and memory. Cold Spring Harbor Perspect Biol 2015;7:a021808.

[152] Stachenfeld KL, Botvinick MM, Gershman SJ. The hippocampus as a predictive map. Nat Neurosci 2017;20:1643–1653.

[153] Whittington JC, Muller TH, Mark S, Chen G, Barry C, Burgess N, Behrens TE. The Tolman-Eichenbaum Machine: Unifying space and relational memory through generalisation in the hippocampal formation. Preprint bioRxiv:770495, 2019.

[154] Lloyd K, Leslie DS. Context-dependent decision-making: A simple Bayesian model. J R Soc Interface 2013;10:20130069.

[155] Fields C, Levin M. Multiscale memory and bioelectric error correction in the cytoplasm-cytoskeleton-membrane system. WIRES Syst. Biol. Med. 2018;10:e1410.

[156] Fields C, Glazebrook JF, Levin M. Neurons as hierarchies of quantum reference frames. Preprint arxiv:2209.00921 [q-bio.NC], 2022.

[157] Kuchling F, Fields C, Levin M. Metacognition as a consequence of competing evolutionary time scales. Entropy 2022;24:601.

[158] Levin, M. Technological approach to mind everywhere: An experimentally-grounded framework for understanding diverse bodies and minds. Front. Syst. Neurosci. 2022;16:768201.

[159] Fields C, Levin M. Competency in navigating arbitrary spaces: Intelligence as an invariant for analyzing cognition in diverse embodiments. Preprint PsyArXiv 87nzu (https://psyarxiv.com/87nzu/; accessed 30 Apr 2022)

[160] Safron A. An Integrated World Modeling Theory (IWMT) of consciousness: Combining integrated information and global neuronal workspace theories with the Free Energy Principle and active inference framework; Toward solving the Hard Problem and characterizing agentic causation. Front Artif Intell 2020;3:30.

[161] Davis MH, Johnsrude IS. Hierarchical processing in spoken language comprehension. J Neurosci 2003;23:3423–3431.

[162] George D, Hawkins J. Towards a mathematical theory of cortical micro-circuits. PLoS Comput Biol 2009;5:e1000532.

[163] Friston KJ, Lin M, Frith CD, Pezzulo G, Hobson JA, Ondobaka S. Active inference, curiosity and insight. Neural Comput 2017;29:2633–2683.

[164] Friston KJ., Rosch, R, Parr T, Price C, Bowman H. Deep temporal models and active inference. Neurosci Biobehav Rev 2017;77:388–402.

[165] MacKay DJ. Free-energy minimisation algorithm for decoding and cryptoanalysis. Electron Lett 1995;31:445–447.

[166] Teh YW, Jordan MI, Beal MJ, Blei DM. Hierarchical Dirichlet processes. J Am Statist Assoc2006;101:1566–1581.

[167] Friston KJ, Parr T, Yufik Y, Sajid N, Price CJ, Holmes E. Generative models, linguistic communication and active inference. Neurosci Biobehav Rev 2020;118:42–64.

[168] Parr T, Friston KJ. The active construction of the visual world. Neuropsychologia 2017;104:92–101.

[169] Mead C. Neuromorphic electronic systems. Proc IEEE 1990;78:1629–1636.

[170] Tang J, Yuan F, Shen X, Wang Z, Rao M, He Y, Sun Y, Li X, Zhang W, Li Y, Gao B, Qian H, Bi G, Song S, Yang J, Wu H. Bridging biological and artificial neural networks with emerging neuromorphic devices: Fundamentals, progress, and challenges. Adv Mater 2019;31:1902761.

[171] Jaynes ET. Information Theory and Statistical Mechanics. Phys Rev (Series II) 1957;106:620–630.

[172] Friston KJ, Wiese W, Hobson JA. Sentience and the origins of consciousness: From Cartesian duality to Markovian monism. Entropy 2020;22:516.

[173] Maturana H, Varela FJ. Autopoiesis and Cognition: The Realization of the Living. Boston Stud Phil Sci 42. D. Reidel, Dordrecht, 1980.

[174] Varela F, Thompson E, Rosch, E. The Embodied Mind: Cognitive Science and Human Experience. MIT Press, Cambridsge, MA, 1991.

[175] Anderson ML. Embodied cognition: A field guide. Artif Intell 2003;149:91–130.

[176] Froese T, Ziemke T. Enactive artificial intelligence: Investigating the systemic organization of life and mind. Artif Intell 2009;173:466–500.

[177] Schrödinger E. What Is Life?. Cambridge University Press, Cambridge, UK, 1944.

[178] Wigner EP. Remarks on the mind-body question. In: Good IJ (ed.), The Scientist Speculates. London: Heinemann, 1961, pp. 284–302.

[179] Penrose R. The Emperor's New Mind. Oxford University Press, Oxford, 1989.

[180] Hameroff S, Penrose R. Orchestrated reduction of quantum coherence in brain microtubules: A model for consciousness. Math. Comput. Simul. 1996;40:453–480.

[181] Tegmark M. Importance of quantum decoherence in brain processes. Phys. Rev. E 2000;61:4194–4206.

[182] Davies PCW. Does quantum mechanics play a non-trivial role in life? BioSystems 2004;78:69–79.

[183] Hameroff S, Tuszynski, J. Quantum states in proteins and protein assemblies: The essence of life? In: Abbott D, Bezrukov SM, Der A, Sánchez A. (eds) Fluctuations and Noise in Biological, Biophysical, and Biomedical Systems II SPIE, Bellingham, WA, 2004, pp. 27–41.

[184] Arndt M, Juffmann T, Vedral V, Quantum physics meets biology. HFSP J 2009;3:386–400.

[185] Lambert N, Chen Y-N, Cheng Y-C, Li C-M, Chen G-Y, Nori F. Quantum biology. Nat. Phys. 2012;9:10–18.

[186] Melkikh AV, Khrennikov A. Nontrivial quantum and quantum-like effects in biosystems: Unsolved questions and paradoxes. Prog. Biophys. Mol. Biol. 2015;119:137–161.

[187] Brookes JC. Quantum effects in biology: Golden rule in enzymes, olfaction, photosynthesis and magnetodetection. Proc. R. Soc. A 2017;473:20160822.

[188] McFadden J, Al-Khalili J. The origins of quantum biology. Proc. R. Soc. A 2018;474:20180674.

[189] Marais A, Adams B, Ringsmuth AK, Ferretti M, Gruber MJ, Hendrikx R, et al. The future of quantum biology. J. R. Soc. Interface 2018;15:20180640.

[190] Cao J, Cogdell RJ, Coker DF, Duan H-G, Kleinekathöfer U, Jansen TLC, et al. Quantum biology revisited. Sci. Adv. 2020;6:eaaz4888.

[191] Kim Y, Bertagna F, D'Souza EM, Heyes DJ, Johannissen LO, Nery ET. Quantum biology: An update and perspective. Quant Rep 2021;3:1–48.

[192] Polanyi M. Life's irreducible structure. Science 1968;160:1308–1312.

[193] Rosen R. On information and complexity. In Casti JL, Karlqvist A. (eds) Complexity, Language, and Life: Mathematical Approaches. Springer, Berlin, 1986; pp. 174–196.

[194] Ellis GFR. Physics, complexity, and causality. Nature 2005;435:743.

[195] Fields C, Levin M. Metabolic limits on classical information processing by biological cells. BioSystems 2021;209:104513.

[196] Cervantes VH, Dzhafarov EN. Snow Queen is evil and beautiful: Experimental evidence for probabilistic contextuality in human choices. Decision 2018;5(3):193–204.

[197] Basieva I, Cervantes VH, Dzhafarov EN, Khrennikov A. True contextuality beats directs influences in human decision making. J Expt Psychol General 2019;148:1925–1937.

[198] Khrennikov A. Quantum-like modeling of cognition. Front Phys 2015;3:77.

[199] Basieva I, Khrennikov A, Ohya M, Yamato O. Quantum-like interference effect in gene expression: glucose-lactose destructive interference. Syst Synth Biol 2011;5:59–68.

[200] Friston K, Frith C. A Duet for one. Conscious Cogn 2015;36:390–405.

[201] Frith C, Wentzer T. Neural Hermeneutics. In Kaldis B. (ed) Encyclopedia of Philosophy and the Social Sciences. SAGE Publications, Thousand Oaks, CA, 2013, pp. 657–659.

[202] Friston KJ, Frith CD. Active inference, communication and hermeneutics. Cortex 2015;68:129–143.

[203] Isomura T, Parr T, Friston K. Bayesian filtering with multiple internal models: Toward a theory of social intelligence. Neural Comput 2019;31:2390–2431.

[204] Campbell JO. Universal Darwinism as a process of Bayesian inference. Front Syst Neurosci 2016;10:49.

[205] Fields C, Levin M. Integrating evolutionary and developmental thinking into a scale-free biology. BioEssays 2020;42:1900228.

[206] Fields C, Levin M. Does evolution have a target morphology? Organisms 2020;4:57–76.

[207] Zurek WH. Quantum Darwinism. Nature Physics 2009;5:181–188.

[208] Campbell JO. Quantum Darwinism as a Darwinian process. Preprint arxiv:1001.0745 [physics.gen-ph], 2010.

[209] Fields C. Quantum Darwinism requires an extra-theoretical assumption of encoding redundancy. Int J Theor Phys 2010;49:2523–2527.

[210] Abramsky S, Brandenburger, A. The sheaf-theoretic structure of non-locality and contextuality. New J Phys 2011;13:113036.

[211] Abramsky S, Hardy L. Logical Bell inequalities. Phys Rev A 2012; 85(6):062114.

[212] Popescu S. Non-locality beyond quantum mechanics. Nat Phys 2014;10:264–270.

[213] Adlam E. Contextuality, fine-tuning and teleological explanation. Preprint arxiv:2110.15898v1[quant-ph], 2021.

**Journal: Progress in Biophysics and Molecular Biology**
**Article Type: Review article Manuscript Number: PBMB-D-22-00024**
**Article: A free energy principle for generic quantum systems**
**Authors: Chris Fields, Karl Friston, James F. Glazebrook, and Michael Levin**

The authors declare that they have no conflicts of interest in the present work.