

## Review

# Information and Efficiency in the Nervous System—A Synthesis

Biswa Sengupta<sup>1,2\*</sup>, Martin B. Stemmler<sup>3</sup>, Karl J. Friston<sup>1</sup>

**1** The Wellcome Trust Centre for Neuroimaging, University College London, London, United Kingdom, **2** Centre for Neuroscience, Indian Institute of Science, Bangalore, India, **3** Bernstein Centre Munich, Institute of Neurobiology, Ludwig Maximilians Universität, München, Germany

**Abstract:** In systems biology, questions concerning the molecular and cellular makeup of an organism are of utmost importance, especially when trying to understand how unreliable components—like genetic circuits, biochemical cascades, and ion channels, among others—enable reliable and adaptive behaviour. The repertoire and speed of biological computations are limited by thermodynamic or metabolic constraints: an example can be found in neurons, where fluctuations in biophysical states limit the information they can encode—with almost 20–60% of the total energy allocated for the brain used for signalling purposes, either via action potentials or by synaptic transmission. Here, we consider the imperatives for neurons to optimise computational and metabolic efficiency, wherein benefits and costs trade-off against each other in the context of self-organised and adaptive behaviour. In particular, we try to link information theoretic (variational) and thermodynamic (Helmholtz) free-energy formulations of neuronal processing and show how they are related in a fundamental way through a complexity minimisation lemma.

## Introduction

The design of engineered and biological systems is influenced by a balance between the energetic costs incurred by their operation and the benefits realised by energy expenditure. This balance is set via trade-offs among various factors, many of which act as constraints. In contrast to engineering systems, it has only been possible recently to experimentally manipulate biological systems—at a cellular level—to study the benefits and costs that interact to determine adaptive fitness [1,2]. One such example is the nervous system, where metabolic energy consumption constrains the design of brains [3]. In this review paper, we start by defining computation and information in thermodynamic terms and then look at neuronal computations via the free-energy principle. We then consider the efficiency of information processing in the nervous system and how the complexity of information processing and metabolic energy consumption act as constraints. The final section tries to integrate these perspectives: In brief, we will argue that the principle of maximum efficiency applies to both information processing and thermodynamics; such that—for a given level of accuracy—statistically and metabolically efficient brains will penalise the use of complex representations and associated commodities like energy.

## Information Is Physical

A widely used term in neuroscience is “neuronal computation”; but what does computation mean? Simply put, any transformation of information can be regarded as computation, while the transfer of information from a source to a receiver is communication [4].

To understand the physical basis of computation, let us reconsider Feynman’s example of a physical system whose information can be read out. The example is intentionally artificial, to keep the physics simple, but has a direct parallel to neuroscience, as we will show at the end. Consider a box that it is filled with an ideal gas containing  $N$  atoms. This occupies a volume  $V_1$ , in which we can ignore forces of attraction or repulsion between the particles. Now suppose that the answer to a question is “yes” if all  $N$  atoms are on the right-hand side of the box, and “no” if they are on the left. We could use a piston to achieve this. By compressing the gas into a smaller volume  $V_2$ , a piston performs the work

$$\delta W = P\delta V \quad (1)$$

Classical thermodynamics tells us that the pressure and volume of an ideal gas are linked such that

$$PV = NkT \quad (2)$$

where  $k$  is Boltzmann’s constant and the temperature  $T$  is assumed constant. The work done on the gas is then:

$$\delta W = \int_{V_1}^{V_2} \frac{NkT}{V} dV = NkT(\ln V_2 - \ln V_1) \quad (3)$$

As we compress the gas, the atoms speed up and attain kinetic energy, hence heating the box. According to the conservation of energy, the work done on the gas is converted to heat. This heat is dissipated to the external environment to keep the temperature constant. This means that the internal energy  $U$  of all the particles remains unchanged, such that the work done by the system or change in *Helmholtz free energy*  $A = U - TS$  reduces to the change in

**Citation:** Sengupta B, Stemmler MB, Friston KJ (2013) Information and Efficiency in the Nervous System—A Synthesis. *PLoS Comput Biol* 9(7): e1003157. doi:10.1371/journal.pcbi.1003157

**Editor:** Olaf Sporns, Indiana University, United States of America

**Received:** February 3, 2013; **Accepted:** June 7, 2013; **Published:** July 25, 2013

**Copyright:** © 2013 Sengupta et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This work is supported by a Wellcome Trust/DBT Early Career fellowship to BS. BS is also grateful to financial support obtained from the ESF, Boehringer Ingelheim Fonds, and the EMBO. MBS is supported via funding from the BMBF. KJF is supported by the Wellcome Trust. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: b.sengupta@ucl.ac.uk

thermodynamic entropy  $S = kH$ , where  $H$  is Shannon entropy:

$$\delta A = -\delta W = \delta U - kT\delta H \Rightarrow \delta H = N(\ln V_2 - \ln V_1) \quad (4)$$

For a single gas particle, with  $V_2 = \frac{1}{2}V_1$  we find that Shannon entropy decreases by  $\ln 2$ . This means that by compressing the gas there are fewer places that the particles can occupy and we are less uncertain about their whereabouts. In short, we have gained information. What have we learned from this exercise? To obtain information—in other words, to reduce entropy or average uncertainty—one has to perform work. More generally, Landauer’s seminal work showed that energy is required when information is erased or deleted via irreversible operations [5,6]. In the context of noise or communication, the deletion of incorrect bits therefore requires the dissipation of energy. This dissipation is decreased at lower temperatures because of reduced thermal noise—lower temperatures facilitate a reduction of energy expenditure.

In the brain, volume changes are not the primary mode of conveying information. Instead, the compartments present in the brain, ranging from synaptic clefts to organelles, maintain a relatively constant volume over several seconds at least. What changes on a short time scale are the numbers of molecules, such as transmitters or ions, in these compartments. If we translate volumes to concentrations  $c_i = N/V_i$ , the change in entropy due to information transfer becomes

$$\delta H = N \ln \left( \frac{V_2}{V_1} \right) \rightarrow \delta H = N \ln \left( \frac{c_1}{c_2} \right) \quad (5)$$

The work is then  $\delta W = NkT \ln \left( \frac{c_1}{c_2} \right)$ . If the molecules are charged, the chemical potential sets up an electrical potential (called the Nernst potential), which is the basis for much of the signalling within the brain. For some molecules, such as  $\text{Na}^+$  and  $\text{K}^+$  ions, the concentration changes during electrical signalling are minuscule relative to the total concentrations of these molecules. By linearising  $\delta W$  in the concentration changes, we can easily compute the energetic cost of neuronal signals [7].

In the examples above, the system remains in thermodynamic equilibrium. Recent progress has been made in describing the relationship between Helmholtz free energy and work when the system is driven far from equilibrium—for example, if the gas was compressed quickly. In this more general setting, the Jarzynski equality states [8]:

$$E[-\delta W] \geq \delta A = -kT \ln E \left[ \exp \left( \frac{\delta W}{kT} \right) \right] \quad (6)$$

where the expectation  $E[\cdot]$  is over an ensemble of paths from the initial to final states. Crucially, the change in Helmholtz free energy (and expected work) does not depend upon the path or the rate at which external parameters (like volume) change. Notice that Equation 4 is a special case of Equation 6, when there is only one (infinitely slow) path.

## Summary

In summary, changing the state of a system necessarily entails a change in Helmholtz free energy that is equivalent to the work done on the system. Under isothermal conditions, this changes the thermodynamic entropy, which can be regarded as the average uncertainty or information we have about the (microscopic) state of the system. So is this sufficient to establish the link between

thermodynamic free energy and information processing? Not really: because the information here is about the (microscopic) state of the system in question. This does not speak to representational information of the sort associated with biological computations or communication: information of this sort reflects how one system represents another. In the next section, we consider a purely information theoretic perspective on computation that invokes free energy and entropy of a fundamentally different sort.

## The Free-Energy Principle

Equation 4 shows how the basic laws of classical thermodynamics connect the Helmholtz free energy of a system to its entropy, where entropy corresponds to the disorder or average uncertainty about its state. In biological systems, there is a natural tendency to resist disorder—at multiple levels of organisation. The maintenance of sensory and physiological states within characteristic bounds is typical of biological systems and usually relies on some sort of regulatory process, i.e., homeostasis [9,10]. Mathematically, this can be expressed by saying that the (sensory) states of biological systems have characteristically low Shannon entropy, where—under ergodic assumptions—Shannon entropy is (almost surely) the long-term average of self information or surprise (see below). An ergodic system has an invariant phase volume [11], which is a necessary condition for an organism to exist—in the sense that it would otherwise transgress phase boundaries and cease to exist [12].

Here, the Shannon entropy plays the same role as thermodynamic entropy but measures the dispersion not over microstates of a thermodynamic (canonical) ensemble, but over some phase functions or macroscopic variables that change with time. These variables can take values that are relatively frequent (low surprise) or infrequent (high surprise). Shannon entropy reflects the average surprise of these variables as they fluctuate over time. By minimising the surprise associated with environmental fluctuations (sensory input), an organism can maintain its physiological states within bounds [13,14].

To evaluate surprise, biological systems need to infer the probability of each sensory fluctuation they encounter. In systems like the brain, these inferences need to be made in the blink of an eye. However, calculating the requisite probabilities can be an intricate and lengthy process, making such computations practically intractable. In 1972, the physicist Richard Feynman came up with a clever trick for calculating probabilities (approximately but very efficiently) using *variational free energy* [15]. The trick is to convert a difficult probability density integration problem into an easy optimisation problem by minimising a free energy bound on the quantity of interest—in our case, the surprise of sensory input. In brief, this entails adjusting probability distributions over the causes of sensory input until they minimise the free energy of sensory input. Notice that we have introduced the notion of causes or hidden states of the world that are responsible for generating sensory samples. Heuristically, this means the system or agent has a model of the world that it uses to evaluate the likelihood or surprise of a sensation. Mathematically, hidden states are fictive variables that are necessary to construct a variational free energy bound on surprise, as we will see next.

Let us assume that self-organising systems like the brain represent their environment probabilistically, in terms of hidden states that cause sensory input. For example, an agent might believe its visual sensations were caused by a bird flying across its field of view. These beliefs can be regarded as real-valued, time-

dependent internal or representational states  $\mu(t) \in R$ . These internal states encode a conditional probability density  $q(\psi|\mu)$  over hidden states in the world  $\psi(t) \in \Psi$ —such as the motion, colour, and size of the bird. The objective is to minimise the surprise  $-\ln p(s|m)$  of sensations  $s(t) \in S$ . Here,  $m$  denotes a model entailed by a system or an agent, and  $p(s|m)$  is the probability of observing a particular state under that model. The model is effectively used to generate hypotheses that explain sensory input in terms of hidden states or representations—such as a bird in flight.

As noted above, minimising surprise directly is an intractable problem, so surprise is replaced with its variational free energy bound [15]. This free energy is a function of sensory and internal states and can now be minimised with respect to the internal states:

$$F(t) = E_q[U(s, \psi)] - H[q(\psi|\mu)] \quad (7)$$

$$= -\ln p(s|m) + D[q(\psi) \| p(\psi|s, m)]$$

Here,  $U(t) = -\ln p(s, \psi|m)$  corresponds to an internal energy under a generative model of the world, described in terms of the density over sensory and hidden states  $p(s, \psi|m)$ . In Equation 7 and throughout  $H[p] = E_p[-\ln p]$  denotes the entropy of a probability distribution. Comparison with Equation 4 explains why  $F(t)$  is called free energy—by analogy with its thermodynamic homologue that is defined as internal energy minus entropy. However, it is important to note that variational free energy is not the Helmholtz free energy in Equation 4—it is a functional of a probability distribution over hidden (fictive) states *encoded* by internal states  $q(\psi|\mu)$ , not the probability distribution over the (physical) internal states. This is why variational free energy pertains to information about hidden states that are represented, not the internal states that represent them. In other words, the variational free energy measures the information represented by internal states, not internal states *per se*. Later, we will try to establish the link between variational and Helmholtz free energies. First, we consider the computational implications of minimising variational free energy.

In short, free energy finesses the evaluation of surprise—where an agent can evaluate free energy fairly easily, given the internal energy or a generative model of its environment. The second equality in Equation 7 says that free energy is always greater than surprise, because the second term (Kullback-Leibler divergence) is nonnegative. This means that when free energy is minimised with respect to the internal states, free energy approximates surprise and the conditional density approximates the posterior density over hidden states:

$$D[q(\psi|\mu) \| p(\psi|s, m)] \approx 0 \Rightarrow q(\psi|\mu) \approx p(\psi|s, m) \quad (8)$$

This is known as approximate Bayesian inference, which becomes exact when the conditional and posterior densities have the same form [16]. Intuitively, minimising free energy renders the conditional density the true posterior density over hidden states, where both are informed by—or conditioned on—sensory information. In Bayesian parlance, a posterior density describes a belief after sampling some data—in contrast to a prior belief that existed before the data were available. Minimising variational free energy can therefore be regarded using sensory evidence to update prior beliefs to approximate posterior beliefs.

How can we place a concept like variational free energy in the context of neuronal computation? This has a long history—originating in Geoffrey Hinton [17,18] and Douglas Hofstadter's [19] work using Ising models for inference in artificial neural

networks. Hinton and colleagues realised that variational free energy was mathematically equivalent to the cost function for inference in a neural network, such as a Hopfield model [20]—the difference between the prediction made by the neural network and what it actually produced as an output, i.e., the prediction error. These ideas were subsequently absorbed into the free-energy principle [21,22], whose key insight was that to reduce the entropy of sensations, the system had to act on the environment. The solution is to assume that both the internal states of the system and its action minimise variational free energy (and implicitly surprise). This dual minimisation maps nicely onto perception and action, where variational free energy can be reduced by optimising internal (representational) states or sensory states through active sensory sampling. This is known as active inference and essentially compels organisms to selectively sample what they expect to sample.

Under certain statistical assumptions, free energy is essentially the difference between the agent's predictions and the actual sensations sampled [22]. Therefore, minimising the free energy is equivalent to reducing prediction error and hence surprise [14]. To minimise free energy or prediction error, the brain can either change its prediction to match sensory input or it can change what it samples to match its predictions [21]. This suggests that the brain is continually making predictions and reevaluating them by comparing inputs with internal predictions to make sense of the world. Is there any empirical evidence that this scheme operates in the nervous system?

Volunteers in a magnetic resonance imaging (MRI) scanner watched two sets of moving dots—one random and the other moving coherently. They showed patterns of distributed brain activation that could only be explained in terms of top-down predictions from deep in the brain to visual centres in the occipital cortex. In other words, top-down predictions from the extrastriate cortex appeared to suppress prediction errors in the striate cortex [23]. Assuming the visual system is a hierarchy of cortical areas, such predictive coding enables predictions about hidden states of the world—like coherent motion—to influence processing at lower levels [23]. Similarly, in the auditory cortex, electroencephalographic signals from higher processing centres change brain activity in lower areas [24]. Using dynamic causal modelling, Garrido *et al.* [24] found that models with top-down connections explained empirical electrophysiological data far better than the models with only bottom-up connections. Garrido *et al.* [24] argued that these neuronal responses were consistent with the brain's attempt to conciliate predictions at one level with those in other levels—in other words, to reduce hierarchical prediction error.

What sort of neuronal architectures mediate this prediction error minimisation—or predictive coding? In mammalian brains, cortical areas are organised hierarchically [25,26], wherein populations of neurons can encode expected states of the world and provide top-down predictions to lower or sensory levels [27,28]. For example, top-down connections from pyramidal neurons in the deeper layers of the cortex are thought to provide predictions to superficial pyramidal populations of a lower area. This enables forward connections from superficial pyramidal neurons to convey prediction errors, creating recurrent dynamics that suppress prediction errors at each level of the cortical hierarchy [29–31]. The precision of these errors can be modulated by neuromodulation [32]. Such rescaling of prediction errors in proportion to their precision is simply a form of gain control [33,34] and may mediate attention. In short, the wetware necessary to minimise free energy appears to be available and is remarkably consistent with its known functional anatomy.

In summary, biological organisms are open self-organising systems that operate far from thermodynamic equilibrium [35]. The free-energy principle suggests that organisms avoid phase transitions by minimising (a variational free energy bound on) the Shannon entropy of their sensory states. But how does one reconcile the need of an animal to survive (by avoiding phase transitions) with its innate tendency to forage or explore? This apparent paradox is resolved by noting that active inference is driven by prior beliefs—and these beliefs can entail exploration. In other words, agents expect to explore and would be surprised if they did not. We will return to the central role of priors in the last section.

## Summary

Perception minimises prediction error by optimising synaptic activity (perceptual inference), synaptic efficacy (learning and memory), and synaptic gain (attention and salience) [14]. In doing so, we form an optimal representation of the sensorium. Such strategies of optimisation are mathematically equivalent to predictive coding [36,37] or, as we will see later, maximising the mutual information between sensations and the responses they evoke [38,39]. In the embodied context of action on the environment, free-energy minimisation can also explain active inference in the exteroceptive domain [40] and homeostasis through minimising interoceptive prediction errors. In short, the idea of free-energy minimisation, stemming from Feynman's beautiful piece of mathematics, allows us to consider perception and action under a general framework—and produce testable hypotheses.

## Information Efficiency

In the previous section, we described how variational free energy is intricately linked to surprise—the free-energy principle tells us that an organism should strive to reduce its prediction error thereby reducing free energy. The connection between free energy and information—although obvious—is seldom commented upon (see Table 1 in [41]). To minimise free energy, the expected prediction error has to be minimised while, at the same time, the entropy of the conditional density is maximised. This is slightly paradoxical because the purpose of free-energy minimisation is to reduce sensory entropy. However, Equation 7 shows that if the entropy of sensory states  $H[p(s|m)]$  is minimised vicariously by minimising free energy over time, then the entropy of the conditional density  $H[q(\psi|\mu)]$  must be maximised at each point in time. This follows from a need to balance accuracy and complexity of the sort seen in Occam's razor. We will return to this in a later section in the context of the principle of maximum entropy [42]. In this section, we focus on information theory as a way of describing the quality of representations and the constraints under which these representations are formed.

We know that all animals process and transmit information to survive and reproduce in an uncertain environment. A principled way to understand such signal processing was absent until Claude Shannon's seminal work on information theory [43]. To understand how messages can be transferred efficiently via telegraphic wires, Shannon derived powerful formalisms that provided fundamental limits on communication [43]. On one hand, information theory allowed optimisation of complicated devices like satellite communication systems. On the other hand, it fitted comfortably with the bounds established by thermodynamics [44]. Some years after its inception, biologists used information theory to study the efficiency of processing in the nervous system. It was realised that efficient representations were permitted by

statistical regularities in the sensorium, i.e., hidden states and their sensory consequences that have low entropy (see [45]). However, the influence of random fluctuations and other constraints prohibit a completely efficient encoding of hidden states in the world.

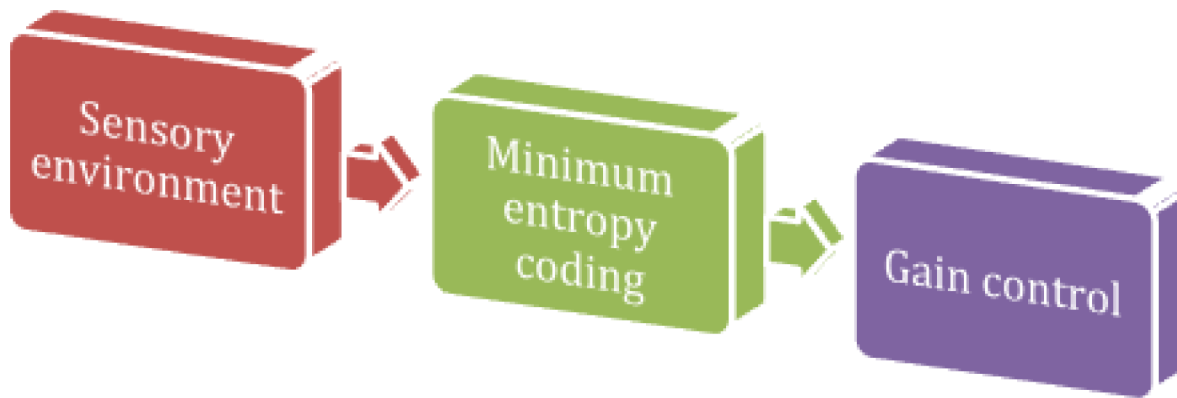
In the nervous system, limited bandwidth and dynamic range create an information bottleneck due to the limited response ranges of the neurons in sensory epithelia [46–48]. Atick [45] suggests that these bottlenecks can also result from computational limitations at higher levels of sensory processing—citing as an example the “attention bottleneck,” where there is constriction of information processing—in bits per unit time—somewhere between area V4 and the inferotemporal cortex. In brief, sensory receptors are required to compress an enormous range of statistically redundant sensory data into their limited range. One way to achieve this is by compression—imagine an architect's plan of your office. This does not include the dimensions of every brick, just the information necessary to build the office. It has been proposed that sensory systems also apply the principle of compression. They sieve redundant information, such that only information that is necessary to encode hidden states is retained [46]—in engineering this is called a factorial code. Of course there are many ways to describe such sensory encoding. Others include but are not restricted to feature detection, filtering, etc. Among these, schemes like linear predictive coding and minimum description length formulations have a particularly close and formal relationship with variational formulations.

Sensory receptors (mechanoreceptors, photoreceptors, and the like) are thought to build a factorial representation of the world—such that only independent bits of information are sampled (Figure 1). Interestingly, this has been observed in the large monopolar cells (LMC) in the blowfly compound eye [49]. Laughlin [49] measured the distribution of the fly's natural environment from horizontal scans of dry woodland and lake-side vegetation and quantified the responses of light-adapted LMCs. Laughlin [49] found that the LMC—known to respond to contrast signals—is most sensitive around the most probable input contrast—with sensitivity dropping to zero as the input became more improbable.

The application of information theory to the nervous system is formally pleasing and has provided some compelling insights. However, it does have limits [50]: although it allows one to quantify the transmission of information, it has no notion of semantics. It only cares about how much information is present but not about what that information represents. A widely used information theoretic metric in neuroscience is the mutual information, which measures how much a random variable tells us about another random variable [51]. If  $s \in S$  is a stimulus and  $\mu \in R$  is the representational response, the mutual information is defined as:

$$I(S; R) = D[p(s, \mu|m) \| p(s|m)p(\mu|m)] \\ = \iint p(s, \mu|m) \ln \frac{p(s, \mu|m)}{p(s|m)p(\mu|m)} ds d\mu \quad (9)$$

Note that the joint density  $p(s, \mu|m)$  is not the generative model  $p(s, \psi|m)$  of the previous section—it describes the joint distribution of sensory and internal states, not the joint distribution of sensory and hidden states. Equation 9 simply describes the divergence or relative entropy between the joint density and the product of its marginals. The mutual information is zero when the neuronal representation is statistically independent of the stimulus and is equal to the entropy of the stimulus when the representation faithfully encodes the stimulus. Since the mutual information must



**Figure 1. Redundancy reduction.** The sensory environment of an animal is highly correlated (redundant). The animal's job is to map such signals as efficiently as possible to its neuronal representations, which are limited by their dynamic range. One way to solve this problem rests on de-correlating the input to provide a minimum entropy description, followed by a gain controller. This form of sensory processing has been observed in the experiments by Laughlin [49], where the circuit maps the de-correlated signal via its cumulative probability distribution to a neuronal response, thereby avoiding saturation. Modified from [45].  
doi:10.1371/journal.pcbi.1003157.g001

lie between zero and channel capacity, it is only the channel capacity that limits the information transfer between stimulus and neuronal response.

Estimating channel capacity by maximising empirical estimates of mutual information can be a difficult task, especially when the experimenter has only an informed guess about the stimuli that evoke responses. One way to finesse this problem is to use adaptive sampling of inputs, which hones in on stimuli that are maximally informative about observed responses [52]. Assuming one knows the stimuli to use, the next problem is the curse of dimensionality. In other words, one requires an enormous amount of data to estimate the probability densities required to quantify mutual information. Although, sophisticated machine learning tools try to estimate mutual information from limited data [53–55], the numerics of mutual information are fraught with difficulties.

## Summary

Irrespective of the thermodynamic or computational imperatives for a biological system, the simple observation that there should be some statistical dependency between sensory samples and the internal states that encode them means that sensory and internal states should have a high mutual information. This leads to the principles of maximum information transfer (a.k.a. infomax) and related principles of minimum redundancy and maximum efficiency [46–48]. Later, we will see how minimising variational free energy maximises mutual information and what this implies for metabolic costs in terms of Helmholtz free energy. First, we will briefly review the biophysical and metabolic constraints on the information processing that underlies active inference.

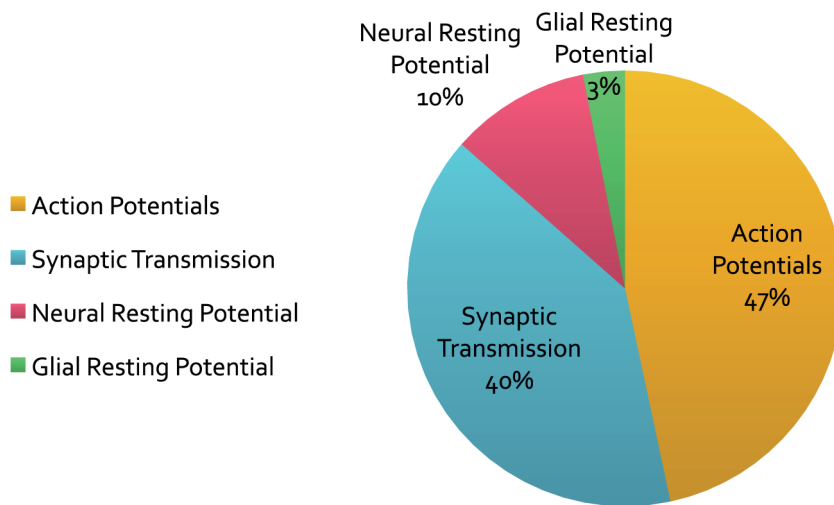
## Is Inference Costly?

Hitherto, we have considered the strategies that neurons might use for abstracting information from the sensorium. A reliable representation is necessary for an animal to make decisions and act. Such information processing comes at a price, irrespective of whether the animal is at rest or not [56]. Cellular respiration enables an organism to liberate the energy stored in the chemical bonds of glucose (via pyruvate)—the energy in glucose is used to produce ATP. Approximately 90% of mammalian oxygen consumption is mitochondrial, of which approximately 20% is uncoupled by the mitochondrial proton leak and 80% is coupled

to ATP synthesis [57]. Cells use ATP for cellular maintenance and signalling purposes, via ion channels that use ATP hydrolysis to transport protons against the electromotive force. Given that the biophysical “cash-register” of a cell (the ATPases) can only handle ATP—and not glucose—we will discuss brain metabolism in terms of ATP.

In man, the brain constitutes just 2% of the body mass, while consuming approximately 20% of the body's energy expenditure for housekeeping functions like protein synthesis, maintenance of membrane potentials, etc. [58]. What consumes such remarkable amounts of energy? Assuming a mean action potential (AP) rate of 4 Hz, a comprehensive breakdown of signalling costs suggests that action potentials use around 47% of the energy consumed—mainly to drive the  $\text{Na}^+/\text{K}^+$  pump (Figure 2) [59]. This pump actively pumps  $\text{Na}^+$  ions out of the neuron and  $\text{K}^+$  ions inside [60]. In doing so, the pump consumes a single ATP molecule for transporting three  $\text{Na}^+$  ions out and two  $\text{K}^+$  ions in [61–63]. Measurements of ATP consumption from intracellular recordings in fly photoreceptors show similar energy consumption to costs obtained from whole retina oxygen consumption [64,65]. Indeed, in the absence of signalling, the dominant cost of maintaining the resting potential is attributable to the  $\text{Na}^+/\text{K}^+$  pump. Attwell and Laughlin [59] further estimated that out of  $3.29 \times 10^9$  ATP/s consumed by a neuron with a mean firing rate of 4 Hz, 47% was distributed for producing APs, while postsynaptic receptors accounted for around 40% of the energy consumption (Figure 2). These figures suggest that action potentials and synapses are the main consumers of energy and that they determine the energy cost in the nervous system.

Experimental studies have shown that neuronal performance is related to energy consumption, both during rest and while signalling [65]. What these studies show is obvious—there is no free lunch. Neurons have to invest metabolic energy to process information. The finite availability of ATP and the heavy demand of neuronal activity suggest neuronal processing has enjoyed great selective pressure. Metabolic energy costs limit not only the possible behavioural repertoire but also the structure and function of many organs, including the brain [3,66,67]. The nervous system can use many tricks to promote energy efficiency. Neurons that use sparse (or factorial) codes for communication [48,68] save on the number of action potentials required to encode information, or



**Figure 2. Attwell and Laughlin's energy budget.** Energy use by various neuronal (cellular) processes that produce, on average, 4 spikes per second. Modified from [59]. doi:10.1371/journal.pcbi.1003157.g002

have topographical connectivity schemes to reduce the surface area of axons connecting different brain areas [69–71]. Neurons may also alter their receptor characteristics to match the probability of inputs to form a matched filter [49]. Alternatively, specialised signal processing could be employed to convert signals from analogue representation to pulsatile—prohibiting accumulation of noise during information transfer [72,73].

In short, nature can use various means to achieve the objective of energy efficiency—see Box 1 for a summary of some strategies. Energy consumption in single neurons depends on the types and the numbers of ion-channels expressed on the lipid bilayer, their kinetics, the cell's size, and the external milieu that changes the equilibrium conditions of the cell. Experimental measures from the blowfly retina show that metabolic efficiency in graded potentials (lacking voltage-gated  $\text{Na}^+$  channels) is at least as expensive as in those neurons displaying action potentials—with the former capable of higher transmission rates [74]. Similarly, in *Drosophila melanogaster* photoreceptors, absence of Shaker  $\text{K}^+$  conductance increases energetic costs by almost two-fold [75,76]. It has also been suggested that the precise mix of synaptic receptors (AMPA, NMDA, mGlu, Kainate, etc.)—that determine synaptic time constants—influences the energetic cost of the single neuron [77]. Recent evidence indicates that the biophysical properties generating an action potential can be matched to make them energy efficient [78–81]. Fast  $\text{Na}^+$  current decay and delayed  $\text{K}^+$  current onset during APs in nonmyelinated mossy fibres in the rat hippocampus minimise the overlap between the inward and outward currents, resulting in a reduction of metabolic costs [81]. Similarly, incomplete  $\text{Na}^+$  channel inactivation in fast-spiking GABAergic neurons during the falling phase of the AP reduces metabolic efficiency of these neurons [78]. Applying numerical optimisation to published data from a disparate range of APs, Sengupta *et al.* [80] showed that there is no direct relationship between size and shape of APs and their energy consumption. This study further established that the temporal profile of the currents underlying APs of some mammalian neurons are nearly perfectly matched to the optimised properties of ionic conductances, so as to minimise the ATP cost. All of these studies show that experimentally measured APs are in fact more efficient than suggested by the previous estimates of Attwell and Laughlin [59]. This was because until 2001 experimental measurements of membrane currents were scant, impeding the study of the overlap between  $\text{Na}^+$  and  $\text{K}^+$

currents. The effects of energy-efficient APs on cortical processing were gauged by recalculating Attwell and Laughlin's (2001) estimates by first using the overlap factor of 1.2—found in mouse cortical pyramidal cells—and then assuming the probability that a synaptic bouton releases a vesicle in response to an incoming spike remains at 0.25 [80]. Neurons that are 80% efficient have two notable effects (Figure 3). First of all, the specific metabolic rate of the cortical grey matter increases by 60%, and second, the balance of expenditure shifts from action potentials to synapses (Figure 3, cf. Figure 2) [80].

The principle of energy efficiency is not just linked to single neurons. Energy budgets have been calculated for the cortex [82], olfactory glomerulus [83], rod photoreceptors [84], cerebellum [85], and CNS white matter [86], among others. These studies highlight the fact that the movement of ions across the cell membrane is a dominant cost, defined by the numbers and cellular makeup of the neurons and the proportion of synaptic machinery embedded in the cell membrane (Figure 4). Niven and Laughlin [3] have argued that when signalling costs are high and resting costs are low, representations will be sparse; such that neurons in a population preferentially represent single nonoverlapping events (also see [87]). Similarly, when resting costs are high and signalling costs are low, the nervous system will favour the formation of denser codes, where greater numbers of neurons within the population are necessary to represent events [3].

Experimental studies of mammalian cortex suggest that the cortex organises itself to minimise total wiring length, while maximising various connectivity metrics [88]. Minimising wiring lengths decreases the surface area of neuronal processes, reducing the energy required for charging the capacitive cell membrane—to sustain and propagate action potentials. In fact, theoretical analyses in pyramidal and Purkinje cells have shown that the dimensions and branching structure of dendritic arbours in these neurons can be explained by minimising the dendritic cost for a potential synaptic connectivity [89,90]. This can result from increasing the repertoire of possible connectivity patterns among different dendrites, while keeping the metabolic cost low [89,90].

## Summary

In summary, we have reviewed several lines of evidence that evolution tries to minimise metabolic costs, where—in the brain—

### Box 1. Some principles of computational anatomy.

**Dimensionality reduction:** Sensory input is high dimensional—a visual scene comprises differences in brightness, colours, numbers of edges, etc. If the retina did not preprocess this visual information, we would have to handle around 36 Gb/s of broadband information, instead of 20 Mb/s of useful data [73]. Preprocessing increases the metabolic efficiency of the brain by about 1,500 times. The requisite dimensionality reduction is closely related to minimising complexity—it is self-evident that internal representations or models of the sensorium that use a small number of dimensions or hidden states will have a lower complexity and incur smaller metabolic costs.

**Energy-efficient signalling:** Action potentials (APs) are expensive commodities, whether they are used for local computation or long-distance communication [59]. Energy-efficient APs are characterised by  $\text{Na}^+$  channel inactivation, voltage-dependent channel kinetics, and cooperative  $\text{K}^+$  channels—as described by multiple gating currents, inward-rectifying  $\text{K}^+$  channels, and high channel densities [7]. These biophysical innovations enable a neuron to produce efficient APs that use the minimal currents necessary to generate a given depolarisation.

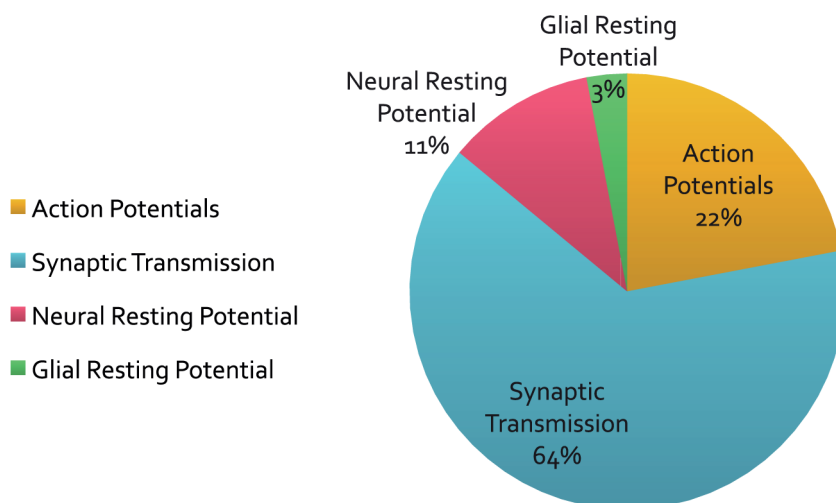
**Component size and numbers:** Action potentials travel considerable distances along densely packed axons, collaterals, and dendrites. The capacitance that must be charged by APs increases with membrane area [101], constraining the number and length of neuronal processes. It is fairly straightforward to show that—to maintain information transfer—the optimal solution is to decrease the number of components. Assuming all neurons have the same thresholds and energy consumption, the energy-efficient solution is to minimise the number of components, under computational constraints dictated by the ecological niche of the animal [101].

**Modular design:** Very-large-scale integration circuits suggest an isometric scaling relation between the number of processing elements and the number of connections (Rent's rule [102]). Neuronal networks have been shown to obey

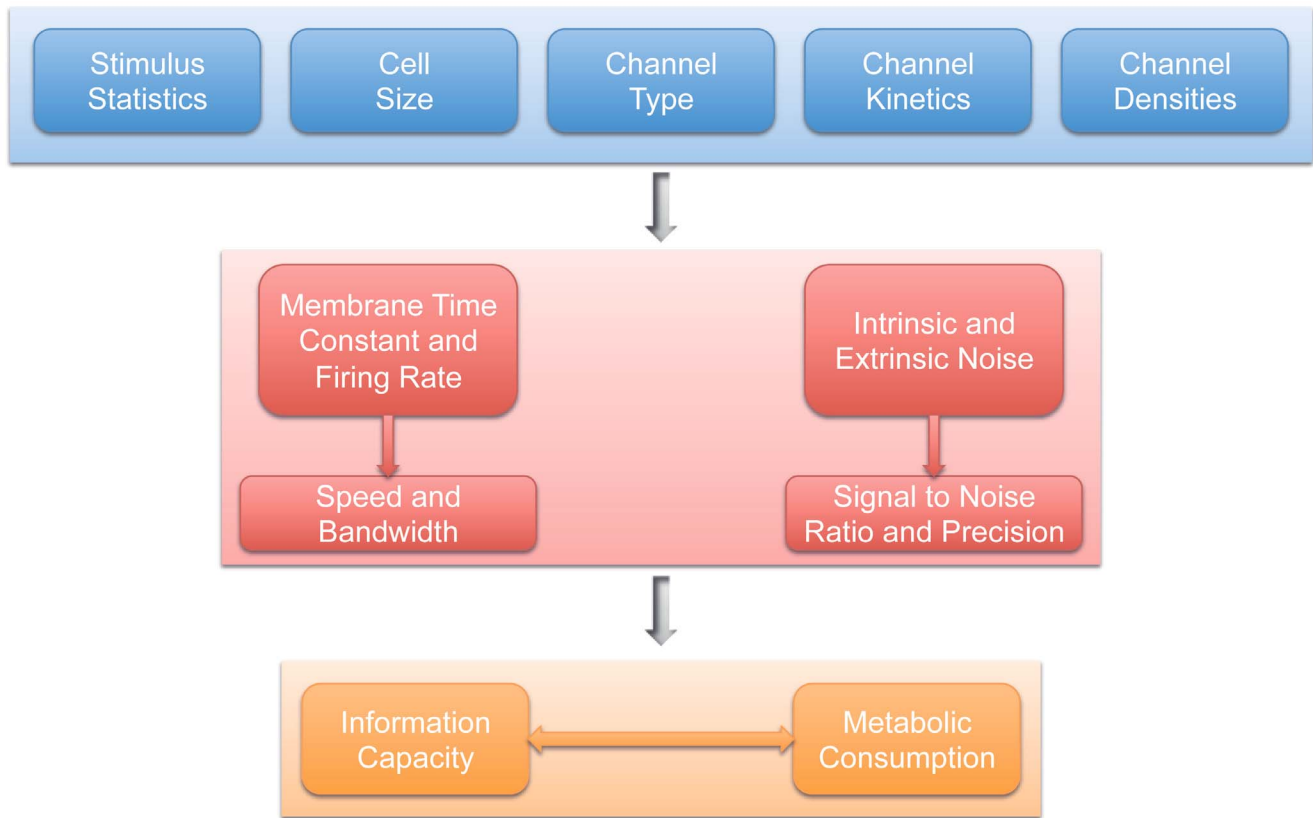
Rent's rule, exhibiting hierarchical modularity that optimises a trade-off between physical cost and topological complexity—wherein these networks are cost-efficiently wired [103]. A modular design balances the savings in metabolic costs, while preserving computational capacities. Hierarchical modularity also emerges under predictive coding [33]. In this context, the brain becomes a model of its environment, which through the separation of temporal scales necessarily requires a hierarchical connectivity.

**Parallel architecture:** The brain processes information in parallel—be it frequency analysis in the inner ear or analysing different attributes of a visual scene using functional segregation. This parallel architecture mirrors those used in modern-day microprocessors. For example, a fast single-core microprocessor may consume 5 Watts and execute a program in 10 seconds. If we bring together two single cores, power will double and execution time will halve, still consuming 50 Joules. Alternatively, a slow double-core microprocessor that expends 2.5 Watts of power to execute the program in 15 seconds could consume only 7.5 Joules. This energy saving works because power is proportional to frequency cubed; therefore, halving the frequency reduces the speed by two but conserves eight times the power, making the microprocessor four times as efficient. In short, if parallel architectures are combined with slow computing speeds, the resulting system is energetically more efficient.

**Analogue versus digital:** If analogue computing is so efficient [104], why don't neurons operate on an all analogue basis? The obvious answer is signal processing in the digital (such as AP) domain enables noise suppression. Noise accumulation in analogue systems [73] speaks to hybrid processing—the use of analogue preprocessing before optimal digitisation. APs are useful in this context because they have an inbuilt threshold mechanism that attenuates noise. If a presynaptic signal is encoded as an AP and transmitted, there is hardly any conduction loss, thereby enabling a reliable transfer of information.



**Figure 3. A revised energy budget for signalling in the grey matter of the rat brain.** Incorporating the increased efficiency of APs in mammalian neurons into Attwell and Laughlin's (Figure 2) original energy budget—for grey matter in the rat brain—reduces the proportion of the energy budget consumed by APs. Modified from [80]. doi:10.1371/journal.pcbi.1003157.g003



**Figure 4. Elements defining metabolic efficiency.** Speed and precision defines the representational capacity of a neuron. Speed or bandwidth is dependent on the membrane time constant and/or the spike rate of the neuron, while precision relies mainly on the types, numbers, and kinetics of synapses and the channels, neuron volume, etc. An efficient brain will maximise speed and precision under energetic constraints.  
doi:10.1371/journal.pcbi.1003157.g004

these costs are primarily incurred by the restoration of transmembrane potentials, whose fluctuations encode or represent hidden states of the world. This raises a question: is energy the only constraint in the evolution of animals? Of course not—functional constraints like reliability, speed, precision, etc. [67] and structural constraints like optimal wiring [91] are equally important. For example, a single action potential in the squid giant axon consumes orders of magnitude more energy than a hippocampal or a pyramidal neuron, yet evolution has invested that extra Joule to buy speed [80,92]. In short, structure and function interact to determine the fitness of an animal. Having surveyed the key metabolic constraints under which neuronal processing must proceed, we now try to integrate the information theoretic and metabolic perspectives.

### Thermodynamic Efficiency and Free-Energy Minimisation

In this section, we gather together the imperatives for biological self-organisation reviewed above. We hope to show that minimising variational free energy necessarily entails a metabolically efficient encoding that is consistent with the principles of minimum redundancy and maximum information transfer. In brief, we will show that maximising mutual information and minimising metabolic costs are two sides of the same coin: by decomposing variational free energy into accuracy and complexity, one can derive the principle of maximum mutual information as a special case of maximising accuracy, while minimising complexity translates into minimising metabolic costs.

### Metabolic Efficiency and Free Energy

To connect the thermodynamic work or metabolic energy required to represent hidden states to the variational free energy of those representations, we need to consider the relationship between representational internal states and the underlying thermodynamic microstates. Recall that internal states  $\mu(t)$  are deterministic quantities that encode a conditional density over hidden states of the world. These macroscopic states can be regarded as *unconstrained internal variables* of a biophysical system; for example, the molar fractions of different molecules in a cellular compartment. The underlying biophysical system can then be associated with a (thermodynamic) canonical ensemble with internal energy:

$$U = \sum_i p_i E(\mu)_i \quad (10)$$

Here,  $p_i$  corresponds to the probability of a particular microscopic state and  $E_i(\mu)$  to its corresponding energy. Given that the total energy is conserved, this probability is given by the Gibbs measure or Boltzmann distribution:

$$p_i = \exp\left(\frac{A - E(\mu)_i}{kT}\right) = \frac{1}{Z} \exp\left(-\frac{E(\mu)_i}{kT}\right) \quad (11)$$

$$A(T, \mu) = -kT \ln Z$$

$$= U - kTH[p_i] = U - TS$$



The partition function  $\mathcal{Z}(T, \mu)$  ensures the probabilities sum to one, while the last equality follows simply from the definition of entropy  $H[p_i] = E_i[-\ln p_i]$ . The Boltzmann distribution describes a system that can exchange energy with a heat bath (or a large number of similar systems) so that its temperature remains constant. The Helmholtz free energy  $A(T, \mu)$  measures the work obtainable from a closed thermodynamic system at a constant temperature and volume—where a closed system can exchange energy with other systems (but not mass).

The key result we will use from statistical thermodynamics is that the Helmholtz free energy is minimised at equilibrium with respect to any unconstrained internal variables for a closed system at constant temperature  $T_0$ ,

$$A_0 = \min_{\mu} A_0(T_0, \mu) \quad (12)$$

where  $A_0(T_0, \mu)$  is the free energy of the system at equilibrium or steady state (i.e., constant entropy). This motivates the following Lemma:

**Lemma:** (complexity minimisation) *Minimising the complexity of a conditional distribution—whose sufficient statistics are (strictly increasing functions of) some unconstrained internal variables of a thermodynamic system—minimises the Helmholtz free energy of that system.*

**Proof:** Using standard results from Bayesian statistics [16], we can express free energy as *complexity* minus *accuracy*

$$\begin{aligned} F(\mu) &= D[q(\psi|\mu) \| p(\psi|m)] - E_q[\ln p(s|\psi, m)] \\ p(\psi|m) &= q(\psi|\mu_0) \end{aligned} \quad (13)$$

The first complexity term is the divergence between the conditional distribution and the prior distribution under the generative model. This effectively counts the degrees of freedom used to encode or predict sensory input. The accuracy is simply the expected log likelihood of the sensory input under the conditional density encoded by internal states. The prior distribution represents beliefs in the absence of sensory input. This corresponds to the distribution encoded by internal states  $\mu = \mu_0$  when deprived of input for a suitably long time—at which point, we can assume thermodynamic equilibrium, such that Helmholtz free energy is minimised (see Equation 12):

$$\begin{aligned} A_0(T_0, \mu_0) &= \min_{\mu} A_0(T_0, \mu) \Rightarrow \\ \mu_0 &= \operatorname{argmin}_{\mu} A_0(T_0, \mu) \end{aligned} \quad (14)$$

However, in the absence of input, variational free energy reduces to complexity  $F_0(\mu) \geq 0$ , which—by Gibbs inequality—has a minimum of zero. This means that complexity is also minimised.

$$\begin{aligned} F_0(\mu) &= D[q(\psi|\mu) \| q(\psi|\mu_0)] \Rightarrow \\ \mu_0 &= \operatorname{argmin}_{\mu} F_0(\mu) \end{aligned} \quad (15)$$

In sum, the internal states encoding prior beliefs about hidden states of the world are those that minimise Helmholtz free energy and the complexity defined by variational free energy.

**Remarks:** All we are saying here is that if a (neuronal) system is deprived of sensory inputs it will obtain thermodynamic equilibrium (or at least a nonequilibrium steady state) and will therefore minimise Helmholtz free energy. This assumes, not implausibly, a constant temperature and volume. Crucially, this is precisely the brain state encoding prior beliefs about sensory input, which means that it is the state of minimum computational

complexity. Heuristically, this means that one can associate the complexity cost of variational free energy with metabolic cost—in the sense that they share the same minimum. Crucially, minimising fluctuations in Helmholtz free energy reduces metabolic work by Equation 6. Interestingly, complexity cost also plays a central role in free-energy formulations of optimal control and economic theory [93,94]. Still *et al.* arrive at the same conclusions by treating the thermodynamic system as having an implicit model of its inputs—allowing them to establish the fundamental equivalence between model inefficiency or complexity and thermodynamic inefficiency [95]. However, both of these compelling treatments consider homologues of Helmholtz free energy—not variational free energy, which is a functional of a probabilistic model (the conditional distribution).

## Computational Efficiency and Free Energy

The complexity minimisation lemma suggests that commonly occurring representational states—that are *a priori* most probable—are the least costly; for example, resting levels of transmembrane voltage or baseline firing rates. Rare excursions from these states are associated with a high metabolic cost. But how does minimising complexity relate to principles of minimum redundancy? Because representations do not change sensory inputs, they are only required to minimise the free energy of the conditional density. Assuming conditional uncertainty is small, the conditional density can be approximated with a point mass at  $\tilde{\mu}(t)$ , such that  $q(\psi) = \delta(\psi - \tilde{\mu})$  and the free energy becomes (from Equation 13)

$$\begin{aligned} F(t) &= -\ln p(s(t)|\mu(t)) - \ln p(\mu(t)|m) \Rightarrow \\ \int dt F(t) &\propto H[p(s|\mu, m)] + H[p(\mu|m)] \end{aligned} \quad (16)$$

The first equality expresses free energy in terms of accuracy and complexity, where the second complexity term just reports the surprise about the conditional representation under prior beliefs. The second equality is the corresponding path integral of free energy (known as free action). Under ergodic assumptions [12,96] this can be expressed as the conditional entropy of sensory input, given the representations and the entropy of the internal states. Equation 11 has two important implications. First, it shows that minimising free energy, at each point in time, is equivalent to minimising free action—by the fundamental lemma of variational calculus. In other words, Equation 11 is just a restatement of the principle of least action. Second, it shows that minimising free energy maximises the accuracy of representations or minimises their conditional uncertainty (entropy) over time. This is simply a restatement of the principle of minimum redundancy or maximum mutual information [97]. This follows because minimising uncertainty about sensory inputs, given internal states, implicitly maximises the mutual information between sensory and internal states (for any given sensations):

$$I(S, R) = H[p(s|m)] - H[p(s|\mu, m)] \quad (17)$$

This suggests that the infomax principle [97] is a special case of the free-energy principle that is obtained when we discount uncertainty and represent sensory input with point estimates of their causes. In this context, high mutual information is assured by maximising accuracy (e.g., minimising prediction error) and prior beliefs are enforced by minimising complexity. Crucially, minimising complexity minimises metabolic cost.

In short, the infomax principle can be understood in terms of the decomposition of free energy into complexity and accuracy:

mutual information or statistical efficiency is optimised when conditional expectations maximise accuracy (or minimise prediction error), while thermodynamic efficiency is assured by minimising complexity. This minimisation ensures that the generative model is not over-parameterized and leads to a parsimonious representation of sensory data that conforms to prior beliefs about their causes. Interestingly, advanced model optimisation techniques use free-energy optimisation to eliminate redundant model parameters [98], suggesting that free-energy optimisation might provide a nice explanation for synaptic pruning and homeostasis in the brain during neurodevelopment [99] and sleep [100]. In developing the link between metabolic and statistical efficiency, we have assumed that internal neuronal states encode hidden states in terms of their most likely value or expectation. Is there any principled reason to assume this form of neuronal code?

### The Maximum Entropy Principle and the Laplace Assumption

Notice from Equation 7 that minimising variational free energy entails maximising the entropy of the conditional density. Intuitively, this is like keeping one's options open when trying to find hypotheses or explanations for sensory input. If we admit an encoding of the conditional density up to second order moments, then the maximum entropy principle [42], implicit in the definition of free energy, requires  $q(\psi|\tilde{\mu}) = \mathcal{N}(\tilde{\mu}, \Sigma)$  to be Gaussian. This is because a Gaussian density has the maximum entropy of all forms that can be specified with two moments. Assuming a Gaussian form is known as the Laplace assumption and enables us to express the entropy of the conditional density in terms of its first moment or expectation. This follows because we can minimise free energy with respect to the conditional covariance as follows:

$$F = U(s, \mu) + \frac{1}{2} \text{tr}(\Sigma \cdot \partial_{\mu\mu} U) - \frac{1}{2} \ln |\Sigma| \Rightarrow \partial_{\Sigma} F = \frac{1}{2} \partial_{\mu\mu} U - \frac{1}{2} \Pi$$

$$\partial_{\Sigma} F = 0 \Rightarrow \begin{cases} \Pi = \partial_{\mu\mu} U \\ F = U(s, \mu) + \frac{1}{2} \ln |\partial_{\mu\mu} U| \end{cases} \quad (18)$$

Here, the conditional precision  $\Pi(\mu)$  is the inverse of the conditional covariance  $\Sigma(\mu)$ . Equation 18 means the free energy becomes a function of conditional expectations and sensory states. This is important because it suggests the brain may represent hidden states of the world in terms of their expected values. This leads to the Laplace code (defined as neuronal encoding under the Laplace assumption), which is arguably the simplest and most flexible of all neuronal codes [13,14]. Furthermore, under the Laplace code, one can minimise free energy efficiently using predictive coding [29,31]. Predictive coding has become one of the most popular ways of understanding message passing in the

brain—particularly in the setting of hierarchical perceptual inference. In short, the free-energy principle entails the principle of maximum entropy and leads, in a principled way, to a neuronal encoding of representations in terms of conditional expectations.

The specific nature of the neural code may be exclusive to a species or underlying neural function. Whatever its makeup—expected latency, firing rate, spike timing, phase, etc.—it will exist to harmonize the dialogue between perception and action. In practice, we usually have in mind the instantaneous rate of firing of neuronal populations, which means the internal states encoding posterior beliefs are ensemble averages of ensemble averages—for example, the expectation of (a function of) depolarisation over the neuronal ensemble, where the depolarisation of a single neuron is (a function of) the internal variables of a canonical ensemble.

### Conclusion

We have reviewed the thermodynamic and computational (statistical) imperatives for biological self-organisation, with a special focus on neuronal circuits. We have considered the role of classical thermodynamics and the notion of metabolic efficiency—that appears to be an important constraint, under which neurophysiology and neuroanatomy have evolved. From a computational perspective, we have looked at variational free-energy minimisation as the basis for active Bayesian inference and modelling of the environment. The ability to represent and predict hidden environmental states efficiently can be quantified in terms of mutual information. Our synthesis suggests that minimising variational free energy is a sufficient account of the tendency to maximise both metabolic and statistical efficiency. The motivation for minimising variational free energy is to minimise its long-term average to maintain a constant external milieu—as measured by the entropy of an organism's sensory samples over time. By decomposing variational free energy into accuracy and complexity one can understand metabolic efficiency in terms of minimising complexity (which minimises Helmholtz free energy), under the computational constraint that sensory inputs are represented accurately. Conversely, statistical efficiency can be understood in terms of maximising the accuracy (which maximises mutual information), under the constraint that representations have minimal complexity. The link between complexity and metabolic cost rests on the simple observation that, in the absence of sensory input, prior beliefs are encoded by physical variables that minimise Helmholtz free energy.

The nice thing about this formulation is that, under active inference, organisms will selectively sample sensory inputs that conform to their prior beliefs and minimise the complexity of their representations. This means that biological systems will appear to act in a way that minimises fluctuations in Helmholtz free energy—and will aspire to the nonequilibrium steady state that has been assigned to them by evolution.

### References

- Sutherland WJ (2005) The best solution. *Nature* 435: 569.
- Alexander RM (1996) *Optima for animals*. Princeton/Chichester: Princeton University Press.
- Niven JE, Laughlin SB (2008) Energy limitation as a selective pressure on the evolution of sensory systems. *J Exp Biol* 211: 1792–1804.
- Feynman RP, Hey AJG, Allen RW (1996) *Feynman lectures on computation*. Reading (Massachusetts): Addison-Wesley.
- Landauer R (1961) Irreversibility and heat generation in the computing process. *IBM Journal of Research and Development* 5: 183–191.
- Landauer R (1996) Minimal energy requirements in communication. *Science* 272: 1914–1918.
- Stemmler M, Sengupta B, Laughlin SB, Niven JE (2011) Energetically optimal action potentials. In: Shew-Taylor J, Zemel RS, Bartlett P, Pereira F, Weinberger KQ, editors. *Advances in neural information processing systems* 24. pp. 1566–1574.
- Jarzynski C (1997) Nonequilibrium equality for free energy differences. *Phys Rev Lett* 78: 2690–2693.
- Ashby WR (1947) Principles of the self-organising dynamic system. *J Gen Psychol* 37: 125–128.
- Nicolis G, Prigogine I (1977) *Self-organisation in non-equilibrium systems*. New York: Wiley.
- Birkhoff GD, Koopman BO (1932) Recent contributions to the ergodic theory. *Proc Natl Acad Sci U S A* 18: 279–282.

12. Friston K, Ao P (2012) Free energy, value, and attractors. *Comput Math Methods Med* 2012: 937860.
13. Friston K (2009) The free-energy principle: a rough guide to the brain? *TICS* 13: 293–301.
14. Friston K (2010) The free-energy principle: a unified brain theory? *Nat Rev Neurosci* 11: 127–138.
15. Feynman RP (1998) *Statistical mechanics: a set of lectures*. Boulder (Colorado): Westview Press.
16. Beal M (2003) Variational algorithms for approximate bayesian inference [PhD thesis]. London: University College London.
17. Hinton GE, Sejnowski TJ (1983) Analyzing cooperative computation. In: *Proceedings of the 5th Annual Congress of the Cognitive Science Society*; 1983; Rochester, New York, United States.
18. Hinton GE, Sejnowski TJ (1983) Optimal perceptual inference. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*; 1983; Washington, D.C., United States. pp. 448–453.
19. Hofstadter DR (1984) *The copycat project: an experiment in nondeterminism and creative analogies*. Boston: MIT Artificial Intelligence Laboratory.
20. Hopfield JJ (1982) Neural networks and physical systems with emergent collective computational abilities. *Proc Natl Acad Sci U S A* 79: 2554–2558.
21. Friston K, Kilner J, Harrison L (2006) A free energy principle for the brain. *J Physiol Paris* 100: 70–87.
22. Friston K, Mattout J, Trujillo-Barreto N, Ashburner J, Penny W (2007) Variational free energy and the Laplace approximation. *Neuroimage* 34: 220–234.
23. Harrison LM, Stephan KE, Rees G, Friston KJ (2007) Extra-classical receptive field effects measured in striate cortex with fMRI. *Neuroimage* 34: 1199–1208.
24. Garrido MI, Kilner JM, Kiebel SJ, Friston KJ (2007) Evoked brain responses are generated by feedback loops. *Proc Natl Acad Sci U S A* 104: 20961–20966.
25. Zeki S, Shipp S (1988) The functional logic of cortical connections. *Nature* 335: 311–317.
26. Felleman DJ, Essen DCV (1991) Distributed hierarchical processing in the primate cerebral cortex. *Cereb Cortex* 1: 1–47.
27. Kawato M, Hayakawa H, Inui T (1993) A forward-inverse optics model of reciprocal connections between visual cortical areas. *Network* 4: 415–422.
28. Srinivasan MV, Laughlin SB, Dubs A (1982) Predictive coding: a fresh view of inhibition in the retina. *Proc R Soc Lond B Biol Sci* 216: 427–459.
29. Mumford D (1992) On the computational architecture of the neocortex. II. The role of cortico-cortical loops. *Biol Cybern* 66: 241–251.
30. Crick F, Koch C (1998) Constraints on cortical and thalamic projections: the no-strong-loops hypothesis. *Nature* 391: 245–250.
31. Rao RP, Ballard DH (1999) Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat Neurosci* 2: 79–87.
32. Yu AJ, Dayan P (2005) Uncertainty, neuromodulation, and attention. *Neuron* 46: 681–692.
33. Friston K (2008) Hierarchical models in the brain. *PLoS Comput Biol* 4: e1000211. doi:10.1371/journal.pcbi.1000211.
34. Abbott LF, Varela JA, Sen K, Nelson SB (1997) Synaptic depression and cortical gain control. *Science* 275: 220–224.
35. Prigogine L, Nicolis G, Babloyantz A (1972) Thermodynamics of evolution. *Physics Today* 25: 38–44.
36. Elias P (1955) Predictive coding-I. *IRE Transactions on Information Theory* 1: 16–24.
37. Elias P (1955) Predictive coding-II. *IRE Transactions on Information Theory* 1: 24–33.
38. Stemmler M, Koch C (1999) How voltage-dependent conductances can adapt to maximize the information encoded by neuronal firing rate. *Nat Neurosci* 2: 521–527.
39. Bell AJ, Sejnowski TJ (1995) An information-maximization approach to blind separation and blind deconvolution. *Neural Comput* 7: 1129–1159.
40. Friston K (2011) What is optimal about motor control? *Neuron* 72: 488–498.
41. Peleg Y, Efraim H, Shental O, Kanter I (2010) Mutual information via thermodynamics: three different approaches. *Journal of Statistical Mechanics: Theory and Experiments* Jan: P01014.
42. Jaynes ET (1957) Information theory and statistical mechanics. *Phys Rev* 106: 620–630.
43. Shannon C (1948) A mathematical theory of communication. *Bell System Technical Journal* 27: 379–423.
44. Jaynes ET (2003) *Probability theory: the logic of science*. Cambridge (United Kingdom): Cambridge University Press.
45. Atick JJ (1992) Could information theory provide an ecological theory of sensory processing? *Network* 3: 213–251.
46. Barlow HB, Kaushal TP, Mitchison GJ (1989) Finding minimum entropy codes. *Neural Computation* 1: 412–423.
47. Barlow HB, Kaushal TP, Hawken M, Parker AJ (1987) Human contrast discrimination and the threshold of cortical neurons. *J Opt Soc Am A* 4: 2366–2371.
48. Barlow HB (1959) *Sensory mechanisms, the reduction of redundancy, and intelligence*. National Physical Laboratory Symposium. Teddington (United Kingdom): H.M. Stationery Office.
49. Laughlin S (1981) A simple coding procedure enhances a neuron's information capacity. *Z Naturforsch C* 36: 910–912.
50. Johnson D (2002) Four top reasons mutual information does not quantify neural information processing. In: *Proceedings of the Annual Computational Neuroscience Meeting*; 2002; Chicago, Illinois, United States.
51. Papoulis A (1984) *Probability, random variables, and stochastic processes*. New York: McGraw-Hill.
52. Benda J, Gollisch T, Machens CK, Herz AV (2007) From response to stimulus: adaptive sampling in sensory physiology. *Curr Opin Neurobiol* 17: 430–436.
53. Treves A, Panzeri S (1995) The upward bias in measures of information derived from limited data samples. *Neural Comput* 7: 399–407.
54. Paninski L (2005) Asymptotic theory of information-theoretic experimental design. *Neural Comput* 17: 1480–1507.
55. Paninski L (2003) Estimation of entropy and mutual information. *Neural Computation* 15: 1191–1253.
56. Ames A (2000) CNS energy metabolism as related to function. *Brain Res Brain Res Rev* 34: 42–68.
57. Rolfe DF, Brown GC (1997) Cellular energy utilization and molecular origin of standard metabolic rate in mammals. *Physiol Rev* 77: 731–758.
58. Clarke D, Sokoloff L (1999) Circulation and energy metabolism of the brain. In: Siegel GJ, Agranoff BW, Albers RW, Fisher SK, Uhler MD, editors. *Basic neurochemistry: molecular, cellular and medical aspects*. Philadelphia: Lippincott-Raven. pp. 637–669.
59. Attwell D, Laughlin SB (2001) An energy budget for signaling in the grey matter of the brain. *J Cereb Blood Flow Metab* 21: 1133–1145.
60. Skou JC (1989) The identification of the sodium-pump as the membrane-bound Na<sup>+</sup>/K<sup>+</sup>-ATPase: a commentary on 'The Influence of Some Cations on an Adenosine Triphosphatase from Peripheral Nerves'. *Biochim Biophys Acta* 1000: 435–438.
61. Skou JC (1957) The influence of some cations on an adenosine triphosphatase from peripheral nerves. *Biochim Biophys Acta* 23: 394–401.
62. Post RL, Hegyvary C, Kume S (1972) Activation by adenosine triphosphate in the phosphorylation kinetics of sodium and potassium ion transport adenosine triphosphatase. *J Biol Chem* 247: 6530–6540.
63. Sen AK, Post RL (1964) Stoichiometry and localization of adenosine triphosphate-dependent sodium and potassium transport in the erythrocyte. *J Biol Chem* 239: 345–352.
64. Pangrsic T, Stusek P, Belusic G, Zupancic G (2005) Light dependence of oxygen consumption by blowfly eyes recorded with a magnetic diver balance. *J Comp Physiol A Neuroethol Sens Neural Behav Physiol* 191: 75–84.
65. Niven JE, Anderson JC, Laughlin SB (2007) Fly photoreceptors demonstrate energy-information trade-offs in neural coding. *PLoS Biol* 5: e116. doi:10.1371/journal.pbio.0050116.
66. Laughlin SB (2001) Energy as a constraint on the coding and processing of sensory information. *Curr Opin Neurobiol* 11: 475–480.
67. Laughlin SB, Sejnowski TJ (2003) *Communication in neuronal networks*. Science 301: 1870–1874.
68. Levy WB, Baxter RA (1996) Energy efficient neural codes. *Neural Comput* 8: 531–543.
69. von der Malsburg C, Willshaw DJ (1977) How to label nerve cells so that they can interconnect in an ordered fashion. *Proc Natl Acad Sci U S A* 74: 5176–5178.
70. Bullmore E, Sporns O (2012) The economy of brain network organization. *Nat Rev Neurosci* 13: 336–349.
71. Bullmore E, Sporns O (2009) Complex brain networks: graph theoretical analysis of structural and functional systems. *Nat Rev Neurosci* 10: 186–198.
72. Sarpeshkar R (1998) Analog versus digital: extrapolating from electronics to neurobiology. *Neural Comp* 10: 1601–1638.
73. Sarpeshkar R (2010) *Ultra low power bioelectronics: fundamentals, biomedical applications, and bio-inspired systems*. Cambridge (United Kingdom): Cambridge University Press.
74. Laughlin SB, de Ruyter van Steveninck RR, Anderson JC (1998) The metabolic cost of neural information. *Nat Neurosci* 1: 36–41.
75. Niven JE, Vähäsöyrinki M, Kauranen M, Hardie RC, Juusola M, et al. (2003) The contribution of Shaker K<sup>+</sup> channels to the information capacity of *Drosophila* photoreceptors. *Nature* 421: 630–634.
76. Niven JE, Vähäsöyrinki M, Juusola M (2003) Shaker K<sup>+</sup> channels are predicted to reduce the metabolic cost of neural information in *Drosophila* photoreceptors. *Proc Biol Sci* 270 Suppl 1: S58–S61.
77. Attwell D, Gibb A (2005) Neuroenergetics and the kinetic design of excitatory synapses. *Nat Rev Neurosci* 6: 841–849.
78. Carter BC, Bean BP (2009) Sodium entry during action potentials of mammalian neurons: incomplete inactivation and reduced metabolic efficiency in fast-spiking neurons. *Neuron* 64: 898–909.
79. Hasenstaub A, Otte S, Callaway E, Sejnowski TJ (2010) Metabolic cost as a unifying principle governing neuronal biophysics. *Proc Natl Acad Sci U S A* 107: 12329–12334.
80. Sengupta B, Stemmler M, Laughlin SB, Niven JE (2010) Action potential energy efficiency varies among neuron types in vertebrates and invertebrates. *PLoS Comput Biol* 6: e1000840. doi:10.1371/journal.pcbi.1000840.
81. Alle H, Roth A, Geiger JR (2009) Energy-efficient action potentials in hippocampal mossy fibers. *Science* 325: 1405–1408.
82. Lennie P (2003) The cost of cortical computation. *Curr Biol* 13: 493–497.
83. Nawroth JC, Greer CA, Chen WR, Laughlin SB, Shepherd GM (2007) An energy budget for the olfactory glomerulus. *J Neurosci* 27: 9790–9800.

84. Okawa H, Sampath AP, Laughlin SB, Fain GL (2008) ATP consumption by mammalian rod photoreceptors in darkness and in light. *Curr Biol* 18: 1917–1921.
85. Howarth C, Peppiatt-Wildman CM, Attwell D (2010) The energy use associated with neural computation in the cerebellum. *J Cereb Blood Flow Metab* 30: 403–414.
86. Harris JJ, Attwell D (2012) The energetics of CNS white matter. *J Neurosci* 32: 356–371.
87. Attneave F (1954) Some informational aspects of visual perception. *Psychol Rev* 61: 183–193.
88. Chklovskii DB, Koulakov AA (2004) Maps in the brain: what can we learn from them? *Annu Rev Neurosci* 27: 369–392.
89. Wen Q, Chklovskii DB (2008) A cost-benefit analysis of neuronal morphology. *J Neurophysiol* 99: 2320–2328.
90. Wen Q, Stepanyants A, Elston GN, Grosberg AY, Chklovskii DB (2009) Maximization of the connectivity repertoire as a statistical principle governing the shapes of dendritic arbors. *Proc Natl Acad Sci U S A* 106: 12536–12541.
91. Chen BL, Hall DH, Chklovskii DB (2006) Wiring optimization can relate neuronal structure and function. *Proc Natl Acad Sci U S A* 103: 4723–4728.
92. Crotty P, Sangrey T, Levy WB (2006) Metabolic energy cost of action potential velocity. *J Neurophysiol* 96: 1237–1246.
93. Ortega PA, Brau DA (2012) Thermodynamics as a theory of decision-making with information processing costs. *ArXiv:1204.6481v1*.
94. Ortega PA, Braun DA (2010) A minimum relative entropy principle for learning and acting. *J Artif Int Res* 38: 475–511.
95. Still S, Sivak DA, Bell AJ, Crooks GE (2012) Thermodynamics of prediction. *Phys Rev Lett* 109: 120604.
96. Birkhoff GD (1931) Proof of the ergodic theorem. *Proc Natl Acad Sci U S A* 17: 656–660.
97. Linsker R (1990) Perceptual neural organization: some approaches based on network models and information theory. *Annu Rev Neurosci* 13: 257–281.
98. Tipping ME (2001) Sparse bayesian learning and the relevance vector machine. *J Mach Learn Res* 1: 211–244.
99. Paus T, Keshavan M, Giedd JN (2008) Why do many psychiatric disorders emerge during adolescence? *Nat Rev Neurosci* 9: 947–957.
100. Gilestro GF, Tononi G, Cirelli C (2009) Widespread changes in synaptic markers as a function of sleep and wakefulness in *Drosophila*. *Science* 324: 109–112.
101. Sengupta B, Faisal AA, Laughlin SB, Niven JE (2013) The effect of cell size and channel density on neuronal information encoding and energy efficiency. *J Cereb Blood Flow Metab*. In press.
102. Christie P, Stroobandt D (2000) The interpretation and application of Rent's rule. *IEEE Journal of VLSI* 8: 639–648.
103. Bassett DS, Greenfield DL, Meyer-Lindenberg A, Weinberger DR, Moore SW, et al. (2010) Efficient physical embedding of topologically complex information processing networks in brains and computer circuits. *PLoS Comput Biol* 6: e1000748. doi:10.1371/journal.pcbi.1000748.
104. Sengupta B (2011) Information encoding and energy consumption in single neuron models [PhD thesis]. Cambridge (United Kingdom): University of Cambridge.