LINK TO ORIGINAL ARTICLE
LINK TO INITIAL CORRESPONDENCE

Is the free-energy principle neurocentric?

Karl Friston

Recently, a free-energy formulation of brain function was reviewed in relation to several other neurobiological theories (The free-energy principle: a unified brain theory? *Nature Rev. Neurosci.* **11**, 127–138 (2010))[1]. Fiorillo raises some interesting questions about the formulation from a neurocentric perspective (A neurocentric approach to Bayesian inference. *Nature Rev. Neurosci.* 14 Jul 2010 (doi: 10.1038/nrn2787-c1))[2]:

> *A primary function of the brain is to infer the state of the world … to determine which motor behaviours will best promote adaptive fitness.*[2]

The free-energy principle generalizes this by assuming that any (biological) system that conserves its form must minimize 'surprise' (maximize adaptive fitness) through exchange with its environment. 'Surprise' is simply the improbability $-\ln p(s|m)$ of sensory data s, given a model m of the environment that is entailed by the form of the system. Exchange with the environment transcends motor behaviour and could cover phototropism in plants (which expect their foliage to be deployed in sunlight) to the elaboration of dendritic processes by a neuron sampling its afferents. In all cases the system tries to sample what it expects, under a model of its world.

> *... the free energy approach is divorced from the biophysical reality of the nervous system*[2].

In fact, the approach is grounded explicitly on imperatives for biophysical systems. Furthermore, its neuronal implementation appeals to large bodies of neurophysiological and anatomical facts that often have to be summarized in tables[3,4] (TABLE 1). The premise of the free-energy principle is that an agent is a model of its world, and this model is determined by the agent's biophysical form and states. Mathematically, minimizing average 'surprise' (also called entropy) then becomes the same as maximizing the evidence $p(s|m)$ for its model (that is, itself).

> *... the brain does not need to perform any processing step to go from information to probabilities and inference*[2].

This assertion overlooks the fact that the mapping between environmental causes and sensory consequences is many-to-one (not bijective). This induces ambiguity — when inferring the causes of sensations[5] — that is resolved with (Bayesian) probabilistic inference[6]. A simple example here is that 1 + 4 and 2 + 3 are both causes of 5. Alternative causes can only be represented probabilistically, with

Table 1 | **Biophysical aspects of the brain that can be explained under a free-energy formulation**

Domain	Features explained	Predictions or motivation
Anatomy and connectivity	The hierarchical deployment of cortical areas; recurrent architectures with functionally asymmetric forward and backward connections	Hierarchical cortical organisation
		Distinct neuronal subpopulations that encode expected states of the world and prediction error
		Extrinsic forward connections convey prediction error (from superficial pyramidal cells) and backward connections mediate predictions (from deep pyramidal cells)
		Functional asymmetries in forwards (linear) and backwards (nonlinear) connections are mandated by nonlinearities in the generative model encoded by backward connections
		Principal cells that elaborate predictions (for example, deep pyramidal cells) may show distinct (low-pass) dynamics relative to those that encode error (for example, superficial pyramidal cells)
		Recurrent dynamics are intrinsically stable because they suppress prediction error (no strong loops)
Synaptic physiology	Both (short-term) neuromodulatory gain-control and (long-term) associative plasticity	Scaling of prediction errors, in proportion to their precision, affords the cortical bias or gain control that is seen in attention
		Short-term modulation of synaptic gain encoding precision or uncertainty (which optimizes a path-integral) must be slower than neuronal dynamics (which optimize free-energy per se)
		Long-term plasticity that is formally identical to Hebbian or associative plasticity
		Neuromodulatory factors may have a dual role in the modulation of postsynaptic responsiveness (for example, through after-hyperpolarizing currents) and synaptic plasticity
Electrophysiology	Classical and extra-classical receptive field effects and long-latency (endogenous) components of evoked cortical responses	Event-related responses are self-limiting transients, where late components rest on top-down suppression of prediction error
		Sensory responses are greater for surprising, unpredictable or incoherent stimuli
		The attenuation of responses that encode prediction error. Together with perceptual learning this explains repetition suppression (for example, mismatch negativity in electroencephalography)
Psychophysiology	The behavioural correlates of some physiological phenomena	For example, priming and global precedence. In cognitive terms, it furnishes a framework in which to model and understand things like perceptual categorisation, temporal sequencing and attention

See REF. 4 for a detailed discussion. Table is reproduced, with permission, from REF. 3 © (2009) Cell Press.

processing that integrates sensory evidence and prior expectations afforded by a (generative) model.

> *Surprise … is essentially just the frequency of an event within an imaginary ensemble of states that could unfold over a long period of time.*[2]

This is a common misconception: surprise (surprisal or self-information) is conditioned on a model and is not an attribute of a sampled (frequentist) distribution. It is $-\ln p(s|m)$ not $-\ln p(s)$. Put simply, surprise depends on predictions, which depend on a model. Agents build models to predict sensations. The model of the world (or the form of an agent) is optimum when it minimizes surprise, at which point the agent's model (or its form) stops changing and is conserved. The free-energy principle is an information-theoretic treatment of systems that conserve themselves over time and is inherently Bayesian.

> *... in apparent contradiction to his hypothesis animals tend to explore the least predictable sensory inputs ...*[2]

Do they? If animals wanted unpredictable sensations they would subject themselves to unprecedented pain. I suspect the deeper question here is how to explain itinerant (wandering or searching) behaviour while minimizing surprise[7]. This is simple to explain: agents use dynamical models (cast mathematically as equations of motion). In other words, agents expect to move through their sensory state-space (because the world is itinerant). Indeed, we have used chaotic exploration to illustrate active inference using free energy[8] (FIG. 1).

> *A truly unified brain theory will need to bridge the gap between Bayesian principles and biophysical reality ...*[2]

Absolutely. Hopefully, these responses affirm that the free-energy principle is fundamentally biocentric in that biophysical states encode probabilistic representations of causal structure in the world and should even apply to single neurons[9].

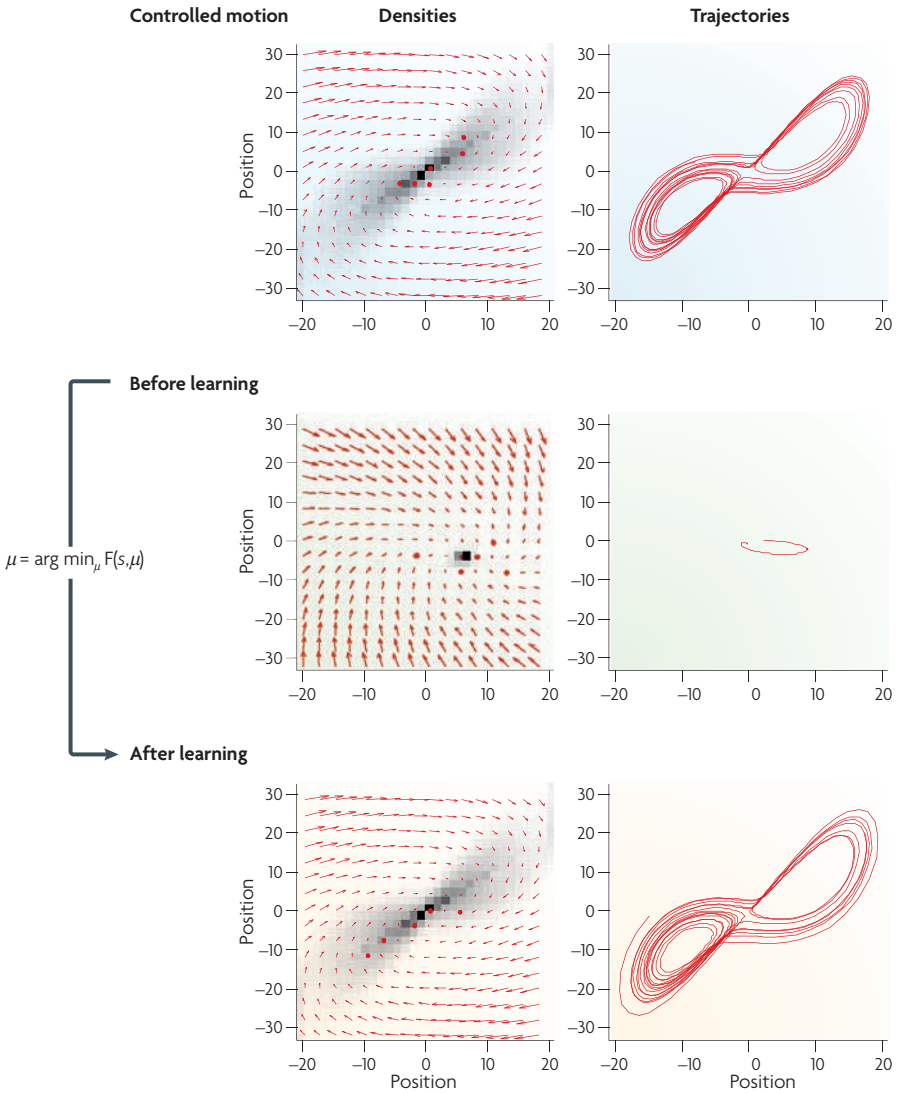

Figure 1 | **The behaviour of an agent that learns to be a Lorenz attractor.** The figure shows the behaviour of an agent that learns to be a Lorenz attractor in terms of equilibrium densities (left) and exemplar trajectories (right). The top panels show the dynamics of a supervised environment that offers control of the agent's motion so that it can experience and learn itinerant (chaotic) behaviour. The middle panels show behaviour before learning, when the agent expects to be drawn to a point attractor. The lower panels show behaviour after learning, when prior expectations about the environment have been transcribed from the environment by learning under the free-energy principle. Here, learning means optimizing the expected parameters (synaptic connection strengths (μ)) of the agent's equations of motion to minimize free-energy F(s, μ). See REF. 8 for details. Figure is reproduced, with permission, from REF. 8 © (2010) Springer.

Karl Friston is at the Wellcome Trust Centre for Neuroimaging, University College London, Queen Square, London WC1N, UK.
e-mail: k.friston@fil.ion.ucl.ac.uk

doi:10.1038/nrn2787-c2
Published online 14 Jul 2010

1. Friston, K. The free-energy principle: a unified brain theory? *Nature Rev. Neurosci.* **11**, 127–138 (2010).
2. Fiorillo, C. D. A neurocentric approach to Bayesian inference *Nature Rev. Neurosci.* 14 Jul (2010) (doi:10.1038/nrn2787-c1).
3. Friston, K. The free-energy principle: a rough guide to the brain? *Trends Cogn. Sci.* **13**, 293–301 (2009).
4. Friston, K. Hierarchical models in the brain. *PLoS Comput. Biol.* **4**, e1000211 (2008).
5. Saberi, K., Takahashi, Y., Farahbod, H. & Konishi, M. Neural bases of an auditory illusion and its elimination in owls. *Nature Neurosci.* **2**, 656–659 (1999).
6. Stone, J. V., Kerrigan, I. S. & Porrill, J. Where is the light? Bayesian perceptual priors for lighting direction. *Proc. Biol. Sci.* **276**, 1797–1804 (2009).
7. Thornton, C. Some puzzles relating to the free-energy principle: comment on Friston. *Trends Cogn. Sci.* **14**, 53–54; author reply 54–55 (2010).
8. Friston, K. J., Daunizeau, J., Kilner, J. & Kiebel, S. J. Action and behavior: a free-energy formulation. *Biol. Cybern.* **102**, 227–260 (2010).
9. Fiorillo, C. D. Towards a general theory of neural computation based on prediction by single neurons. *PLoS ONE* **3**, e3298 (2008).