

# Some puzzles relating to the free-energy principle: Comment on Friston

*Chris Thornton*  
COGS/Informatics  
University of Sussex  
Brighton  
BN1 9QH  
UK  
c.thornton@sussex.ac.uk

January 11, 2010

In Friston's recent article, 'The free-energy principle: a rough guide to the brain' [Friston, Guide, 2009], the structure of an agent's world is taken to be represented by a *conditional density*, a probabilistic mapping from 'causes' to sensory stimulation. Friston argues the brain can arrive at an approximation of this mapping by minimizing *free-energy*, which is a function of sensory stimulation and brain states. A generative model of causal structure in the environment is then obtained, on which basis the agent is able to infer the 'causes of sensory samples' [ibid., p. 294]. What is unclear is how this mechanism would function where sensory samples are ambiguous. In general, there are multiple interpretations for the causes of any sensory data, and these cannot be resolved on the basis of inspecting the data alone [Goldstein, 1996].

For any sense data, there will also generally be causes at multiple levels of description, with causes at one level of description being embedded in causes at higher levels. Sensory stimulation is the result not of distinct causes but of causal structure. How would a mechanism that acts to infer causes measure up to the task of inferring causal structure?

Friston asserts that almost 'any adaptive change in the brain' can be viewed as resulting from minimization of free-energy (p. 293). On the face of it, no particular stand is taken on emergence of the structures that mediate minimization. But looking at the definition of free-energy [Friston, Stephan, 2007], we find a significant role being played by the variable  $\vartheta$ . It is values of this variable that encapsulate the brain's representation of 'environmental causes' [ibid., p. 424]. The range of this variable then dictates the gross structural form of any representation acquired. With the framework providing no principle for deciding the range of  $\vartheta$ , the brain's representation of the conditional density is inevitably a 'slightly mysterious construct' [Friston, Daunizeau, 2009, p. 2].

The expectation may be that  $\vartheta$  will be fixed through instantiation of fortuitous ‘matches’ between internal and external structures. ‘Those systems that can match their internal structure to the causal structure of the environment will attain a tighter bound.’ [Friston and Stephan, 2007, p. 427]. But there is a problem of circularity here — agents are posited to be able to form internal structure matching the environment just in case they already have it. And it is not clear whether this is really intended to be the ‘mechanism’ for fixing  $\vartheta$ . But if there is no principle deciding this critical designator of representational capacity, we can only assume it is fixed at random.

It seems entirely right for Friston to emphasize that the entropic basis of surprise reveals a deep connection between processes of knowledge, behavior and life. But this idea has been in common currency for some years, e.g. [Gatlin 1972; Campbell, 1984], and it is unclear how introduction of the ‘free-energy’ concept, specifically, adds explanatory content. Free-energy is taken to be a ‘good proxy’ for surprise: surely it is minimization of *surprise* that is explanatorily salient. The inability of the present formulation to address the issue of structure-emergence also poses difficulties with regard to specification of  $\vartheta$  ranges, resolution of sensory ambiguity and inference of causal structure.