

Supplementary information S3 (box): The free-energy principle and infomax

Here, we show that the free-energy principle is a probabilistic generalisation of the infomax principle. The infomax principle requires the mutual information $I(\tilde{s}, \mu)$ between sensory data and their conditional representation $\mu(t)$ to be maximal, under prior constraints on the representations; e.g., $p(\mu) = \mathcal{N}(0, I)$. This can be stated as an optimisation of an infomax criterion

$$\begin{aligned} \mu^* &= \arg \max_{\mu} G \\ G &= I(\tilde{s}, \mu) - H(\mu) \\ &= H(\tilde{s}) - H(\tilde{s} | \mu) - H(\mu) \end{aligned} \tag{S3.1}$$

Because the representations do not change sensory data, they are only required to minimise the average surprise about them, given the representations; and the average surprise about the representations, given their prior constraints. These are the last two terms in (S3.1). If the recognition density is a point mass at $\mu(t)$; i.e., $q(\vartheta) = \delta(\vartheta - \mu)$, the free-energy from (S2.4) reduces to

$$F = -\ln p(\tilde{s} | \mu) - \ln p(\mu) \tag{S3.2}$$

From (S1.1), the path-integral of free-energy (also known as free-action) becomes

$$AF = \int dt \quad (\tilde{s}(t), \mu(t)) \mu \quad H(\tilde{s} | \mu) + H(\mu) \tag{S3.3}$$

This means optimising the conditional expectations with respect to free-energy and (by the fundamental lemma of variational calculus) free-action, is exactly the same as optimising the infomax criterion

$$\mu^* = \arg \min_{\mu} F \mathbf{A} = \arg \min_{\mu} \quad = \arg \max_{\mu} G \quad \text{S3.4}$$

In short, the infomax principle is a special case of the free-energy principle that obtains when we discount uncertainty and represent sensory data with point estimates of their causes. Alternatively, the free-energy is a generalisation of the infomax principle that covers probability densities on the unknown causes of data. In this context, high mutual information is assured by maximising accuracy (e.g., minimising prediction error) and the prior constraints are enforced by minimising complexity (see S2.4)