

Supplementary information S4 (box): Value and surprise

Here, we compare and contrast optimal control and free-energy formulations of dynamics on hidden or sensory states. To keep things simple, we will assume the hidden states are known (as is usually assumed in control theory) and ignore random fluctuations; i.e., $\tilde{w}(t) = 0$ (see box S1). In optimum control, one starts with a loss or cost-function (negative reward or utility), $c(\tilde{x})$ and optimises the motion of states to maximise value or expected reward over time

$$\begin{aligned} a^* &= \arg \max_a f(\tilde{x}, a) \cdot \nabla V(\tilde{x}) \\ V(\tilde{x}(0)) &= \int_0^\infty -c(\tilde{x}(t)) dt \Rightarrow \dot{V}(\tilde{x}(t)) = c(\tilde{x}) \end{aligned} \quad \text{S4.1}$$

The first equality says that motion ascends the gradients of the value-function and the second just defines value as reward that will be accumulated in the future. Note the equations of motion $\dot{\tilde{x}} = f(\tilde{x}, a)$ now include action. The value-function is the solution to the celebrated Hamilton-Jacobi-Bellman equation

$$\begin{aligned} \max_a \{ \dot{V}(\tilde{x}(t)) - c(\tilde{x}) \} &= 0 \Rightarrow \\ \max_a \{ f(\tilde{x}, a) \cdot \nabla V(\tilde{x}) - c(\tilde{x}) \} &= 0 \end{aligned} \quad \text{S4.2}$$

This solution ensures that the rate of change of value is cost, as required by the definition of value. In summary, (S4.1) says that action maximises value and (S4.2) means that value is the reward expected under this policy. This ensures low-cost regions attract all trajectories through state-space.

We now revisit value from the perspective of surprise and free-energy. If we put the random fluctuations back and assume a general form (the Helmholtz decomposition) for motion:

$f = \nabla V + \nabla \times W$, it is fairly easy to relate value and surprise (using the Fokker-Planck equation, subject to $\nabla V \cdot (\nabla \times W) = 0$)

$$\begin{aligned} V(\tilde{x}) &= \gamma \ln p(\tilde{x} | m) \\ c(\tilde{x}) &= f \cdot \nabla V + \gamma \nabla^2 V \end{aligned} \quad \text{S4.3}$$

Here, $\gamma > 0$ encodes the amplitude of the random fluctuations (and is known as an inverse sensitivity or temperature parameter). The first equality shows that value is inversely proportional to surprise, where free-energy is surprise because we know the true states. This means the value of a state is proportional to the log-probability of finding an agent m in that state. This is also the log-sojourn time or the proportion of time the state is occupied by that agent.

In the limit of small fluctuations $\gamma \rightarrow 0$, the ensemble density $p(\tilde{x} | m) = \exp(\gamma^{-1} V(\tilde{x}))$ becomes a point mass at the minimum of the cost-function. This somewhat trivial case serves to connect optimal control theory to the equilibrium treatment that underpins the free-energy scheme. In this limit, cost is just the rate of change of value: $c(\tilde{x}) = f \cdot \nabla V = \dot{V}(\tilde{x}(t))$, as mandated by the definition of value in Equation S4.1, which is the solution to the (deterministic) Hamilton-Jacobi-Bellman equation (S4.2).

Crucially, Equation S4.3 also shows that peaks of the equilibrium density can only exist where cost is zero or less

$$\left. \begin{aligned} \nabla V(x) &= 0 \\ \nabla^2 V(x) &\leq 0 \end{aligned} \right\} \Rightarrow c(x) \leq 0 \quad \text{S4.4}$$

with $c(x) = 0$ in the limit $\gamma \rightarrow 0$.

In summary, optimal control theory starts with a cost-function and solves for a value-function that guides the flow or policy to minimise expected cost. Conversely, the equilibrium

perspective starts with flow and derives the implicit value and cost-functions, where value is inversely proportional to surprise. In the last supplementary information box (S5), we show how cost can define policies, without solving the (generally intractable) Hamilton-Jacobi-Bellman equation.