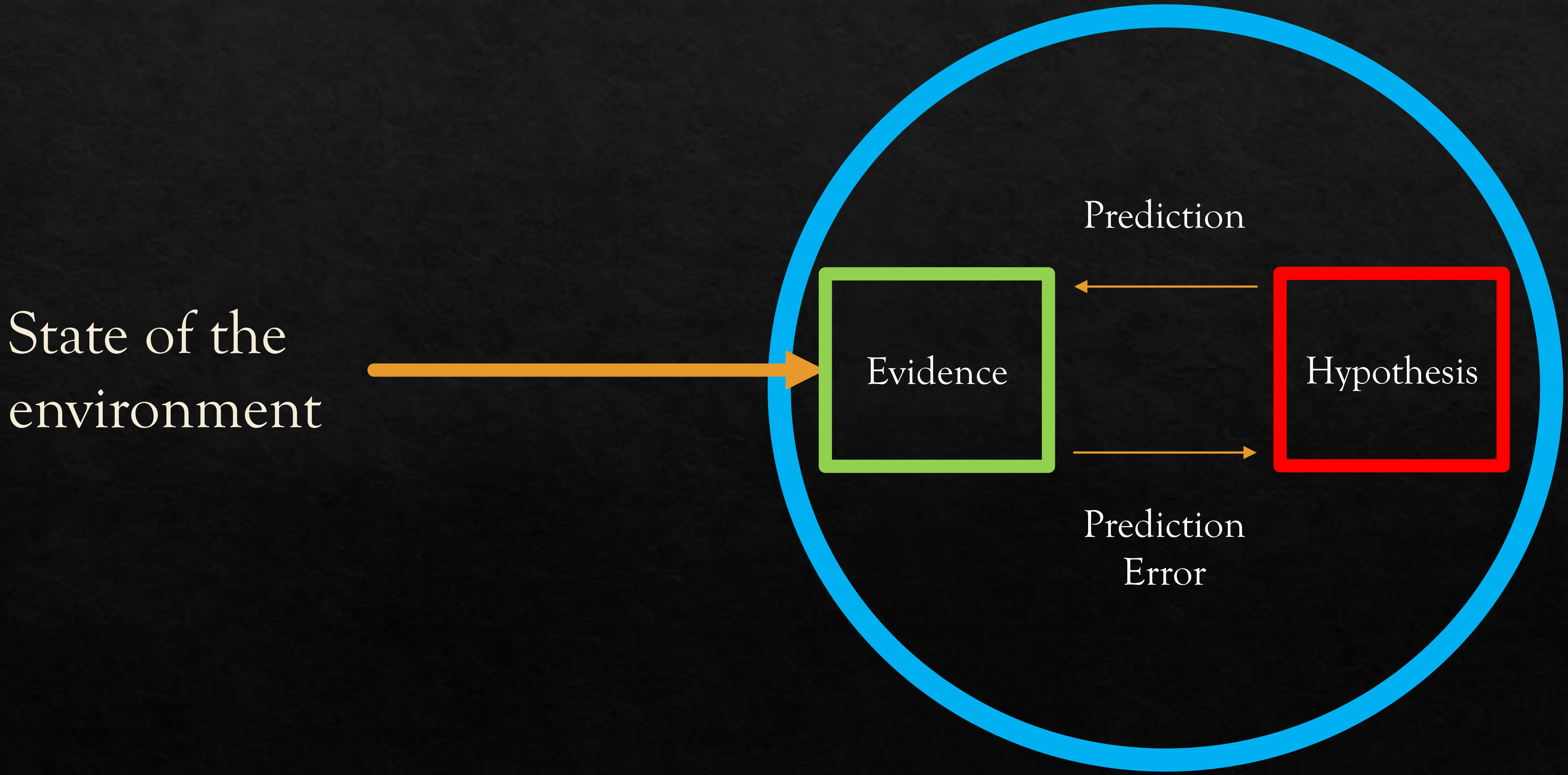
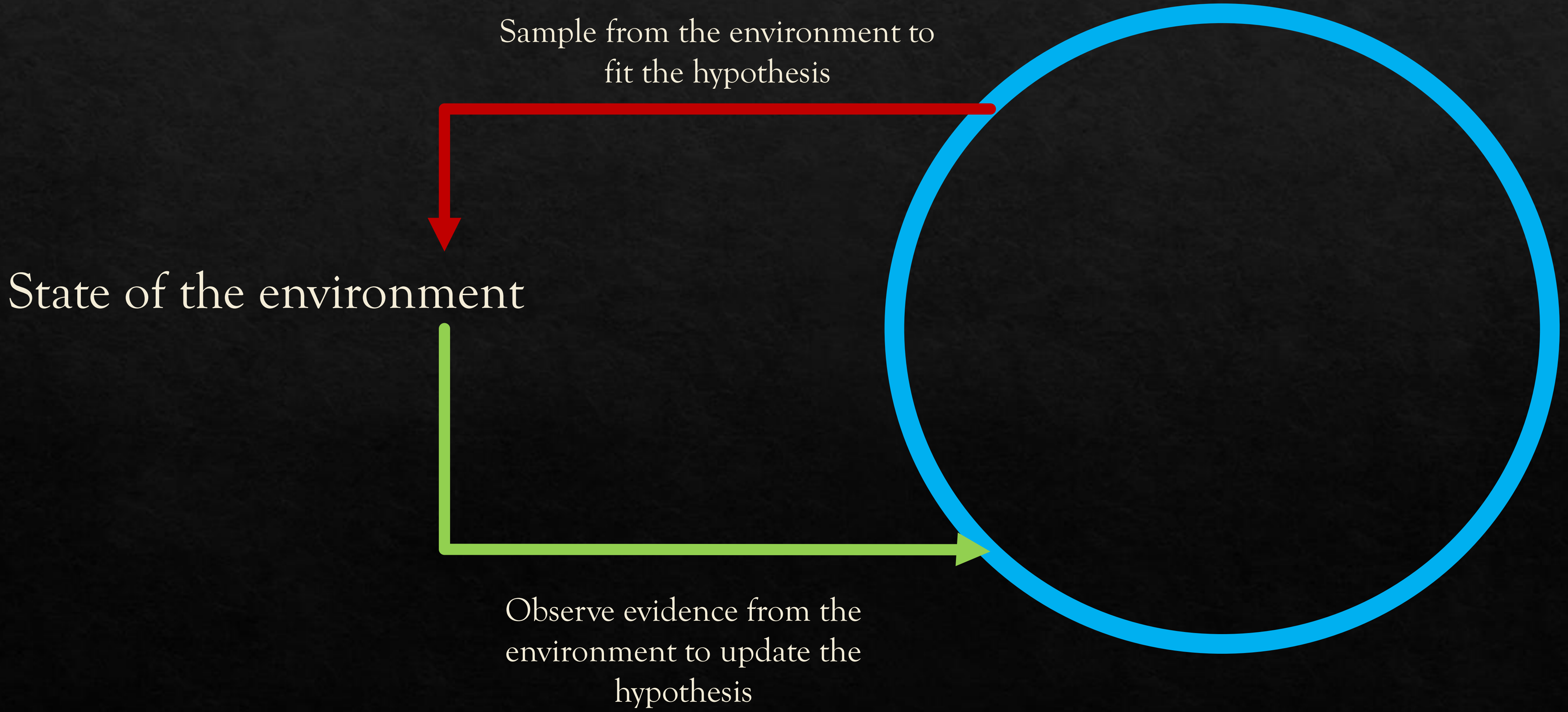
Modeling with Active Inference

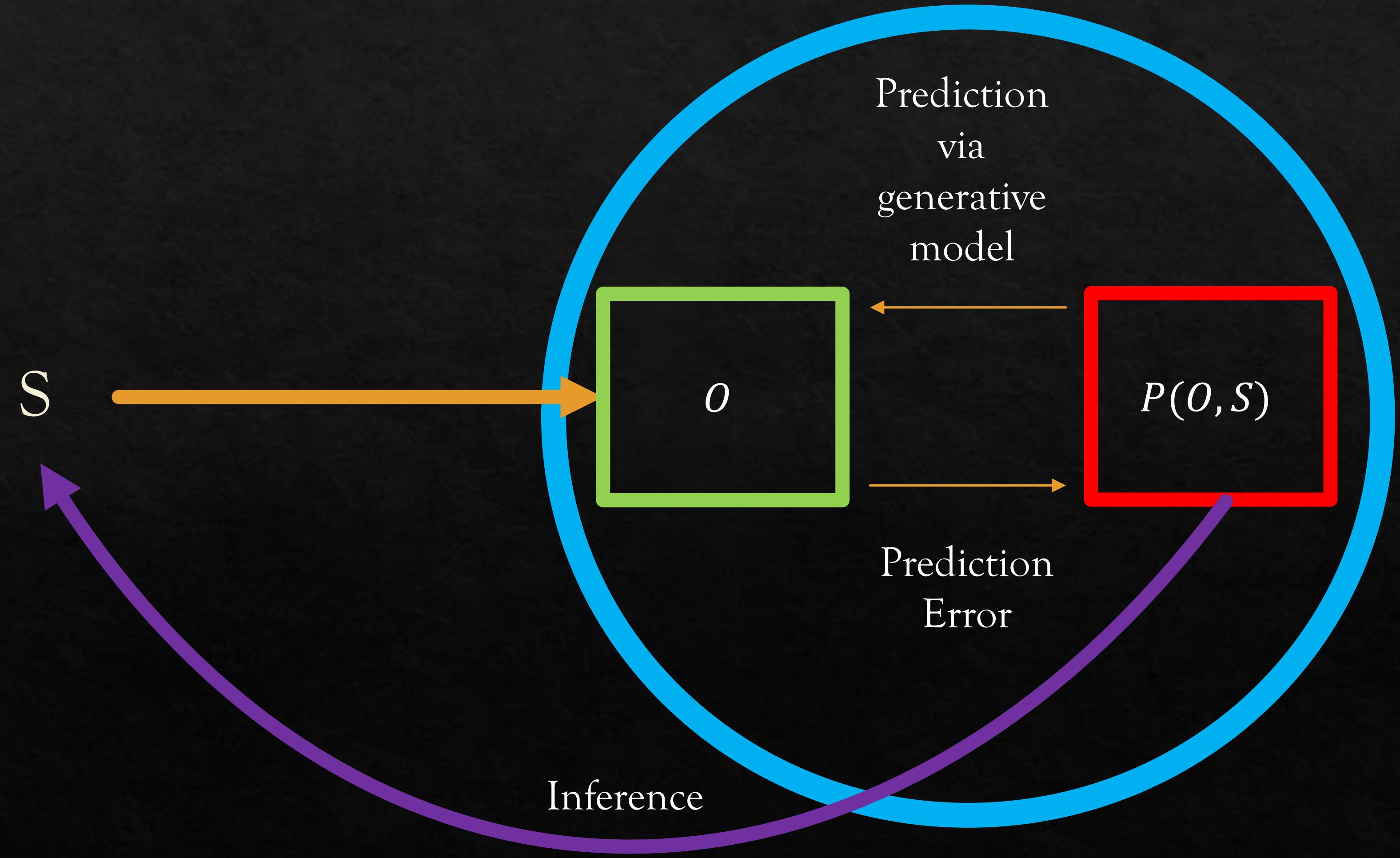
The Brain as an Inference Machine



Perception-Action Cycle



The Brain as a Bayesian Machine



Sometimes an impossible calculation!

$$P(S|O) = \frac{\boxed{P(O|S)P(S)}}{\boxed{\sum_s P(O|S)\,P(S)}}$$

- Easily computable
- Often an intractable calculation

Variational Inference

- A technique is needed to approximate the posterior, due to the intractable denominator.
- As in most variational techniques, an approximating variational posterior (q) is introduced, with the KL-divergence between this variational posterior and the true posterior used to measure their similarity.

$$\sum_{s} q(s) \ln \frac{q(s)}{p(s|o)}$$

- We need to determine some expression, which our crab has control over which can be used to approximate its location

Introducing Free Energy

- One form of the free energy equation can be described as the KL-divergence between an approximating posterior *q* and the generative model, for all states:

$$F \stackrel{\text{def}}{=} \sum_s q(s) \ln \frac{q(s)}{p(o,s)}$$

- With some rearranging:

$$F = \sum_s q(s) \ln \frac{q(s)}{p(s|o)p(o)}$$

$$= \sum_s q(s) \ln \frac{q(s)}{p(s|o)} - \ln p(o)$$

- We now have an easier way to approximate the true posterior: Minimise F!

Model Evidence
(increasing it in terms of perception)

- By minimising the difference between $q(s)$ and $p(s|o)$ we also implicitly increase model evidence, $p(o)$. This is because $p(\text{o}) = \sum_s p(o|s)p(s)$, and after we have approximated the posterior, this becomes our new prior:

$$q(s) \sim = p(s|o)$$

$$p(s) = q(s)$$

$$p(o) = \sum_s p(o|s)p(s)$$

- Conceptually, what is happening here, is that we are iteratively updating part of our model, p(s), to be more congruent with the observation.

Policy Approximation

- With the introduction of an actual agent making decisions and solving a POMDP, it is necessary to include policies in the formulation:

$$F = \sum_{s,\pi} q(s,\pi) ln \frac{q(s,\pi)}{p(s,\pi|o)} - lnp(o)$$

- This formulation makes it clear that, along with state approximation, an agent is also trying to approximate a posterior over policies. Ultimately weighting each policy by how likely it is to produce the observation the agent receives.

- This formulation can be conditioned on a policy: Conceptually, this can be viewed as determining how well a specific policy explains an observation. The policy becomes part of the model, with observations giving evidence to different models.

$$F_\pi = \sum_s q(s|\pi) ln \frac{q(s|\pi)}{p(s|o,\pi)} - \ln p(o)$$

Adding time into the equation

- Within the context of an actual simulation, Free Energy needs to also be a function of time:

$$F(\tau,\pi) = E_{q(s_\tau|\pi)Q(s_{\tau-1}|\pi)} \frac{q(s_\tau|\pi)}{p(o_\tau, s_\tau|s_{\tau-1}, \pi)}$$

- With the free energy for a policy, as a whole expressed as:

$$F(\pi) = \sum_\tau F(\tau,\pi)$$

The Action in Active Inference

- Because the Free Energy equation uses actual observations received by the agent, it can only operate on current and past time steps.
- Expected Free Energy is the free energy calculation at future time steps.
- A crucial difference being, it uses observations as random variables, as it has not actually received them yet.

$$G(\pi) = \sum_{s,o} q(o, s|\pi)\,[lnq(s|\pi) - lnp(o, s|\pi)]$$

- Where $q(o, s|\pi) = p(o|s)q(s|\pi)$

The Preference Distribution

- Another key component introduced with expected free energy is to condition (explicitly or implicitly) the joint probability of states and observations in the generative model upon some desired observations, C.

$$G(\pi) = \sum_{s,o} q(o, s|\pi)\,[lnq(s|\pi) - lnp(o, s|C)]$$

- With some rearranging:

$$G(\pi) = \boxed{\sum_{s,o} q(o, s|\pi)\,[lnp(s|o, \pi) - lnq(s|\pi)]} - \boxed{\sum_{s} q(o|\pi)\,[lnp(o|C)]}$$

- Epistemic
- pragmatic

Model Evidence
(increasing it in terms of action)

- By preferring policies which are expected to produce observations which the agent expects under its model (preferences), the agent is using another means by which to increase model evidence.
- This represents the action component in the perception-action cycle.
- The agent is effectively using action to self-evidence its own hypotheses.

Policy weightings

- Along with state inference, the main task for an agent solving a POMDP is to determine an 'optimal' policy.
- In terms of an active inference agent, this is a policy which is most likely, given both the Free Energy for each policy, the Expected Free Energy and some prior over policies

$$\pi = \sigma(E \; - \gamma G \; - F)$$

Learning the Generative Model

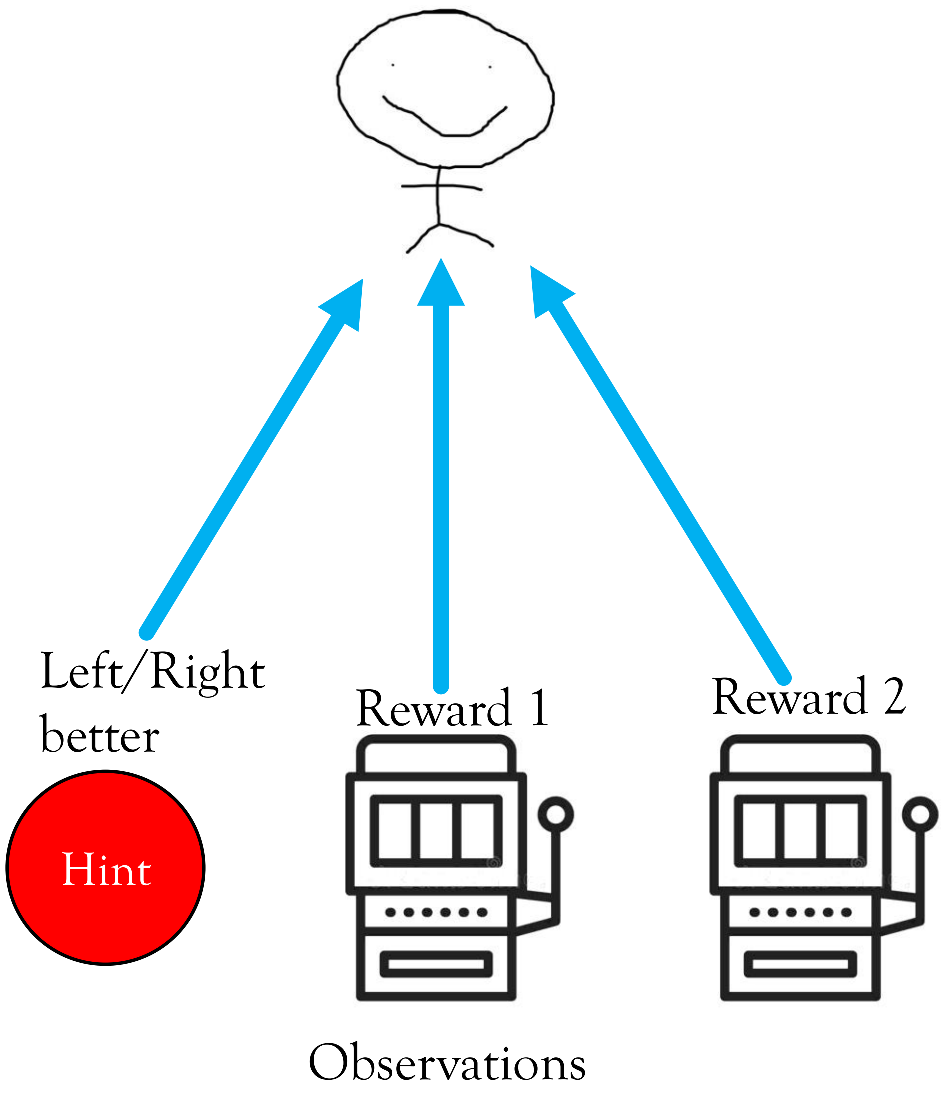
- Crucial to an agent's ability to perform accurate inference is its capacity to have a good likelihood function.
- In the algorithmic implementation of active inference, this can be done quite simply:

$$G(\pi) = D_{kl}[q(o)||p(o|C)] + \sum_{s} q(s|\pi)\,[H[p(o|s)]] - \sum_{s,o} q(o,s|\pi)[D_{kl}[q(\boldsymbol{A}|o,s)||q(\boldsymbol{A})]]$$

where $q(\boldsymbol{A}) = Dir(\boldsymbol{a})$

and $a = p(o|s)$

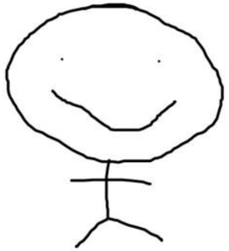
$$\boldsymbol{a}_{trial} = \boldsymbol{a}_{trial-1} + \boldsymbol{n} \times \sum_{\boldsymbol{\tau}} \boldsymbol{o}_{\boldsymbol{t}} \otimes \boldsymbol{s}_{\boldsymbol{\tau}}$$



- Mr Stick can pick the left machine, right machine, or choose the hint.
- Correct choice of the machine = 4 points.
- Incorrect choice = 0 points
- Taking the hint and then picking a machine will only give 2 points

- What is the true state: Left or Right better?
- What hypothesis can Mr Stick infer given an observation

Mr Stick's Generative Model



Likelihood

States

Observations

	Right Better	Left Better
Right Win	$p(R_w \mid R_B)$	$p(R_w \mid L_B)$
Left Win	$p(L_w \mid R_B)$	$p(L_w \mid L_B)$
Right Hint	$p(R_H \mid R_B)$	$p(R_H \mid L_B)$
Left Hint	$p(L_H \mid R_B)$	$p(L_H \mid L_B)$

Priors

Right Better	Left Better
$p(R_B)$	$p(L_B)$