The graphical brain: belief propagation and active inference

Structure of paper

- explain that the brain does Bayesian inference
- consider generative models
- develop a neural network microcircuitry that could implement this belief propagation
- deep hierarchical models
- continuous state-space models in generalised coordinates
- mixed models

Variational free energy

Surprise of an outcome $-\log P(o)$

Free-energy:

Upper bound to surprise

Functional of beliefs

$$F(Q) = E_Q[\log Q(s) - \log P(o, s)]$$

Q is the variational distribution

Expected free energy

Source: active inference- a process theory

$$F(Q) = E_Q[\log Q(s) - \log P(o, s)]$$

$$G(\pi) = \sum_{\tau > t} G(\pi, \tau)$$

$$G(\pi, \tau) = \mathbb{E}_{\tilde{Q}}[\log Q(s_\tau | \pi) - \log P(s_\tau, o_\tau | o_{t<\tau}, \pi)]$$

$$\tilde{Q} := Q(o_\tau, s_\tau | \pi) = P(o_\tau | s_\tau) Q(s_\tau | \pi)$$

$$Q(o_\tau | s_\tau, \pi) = P(o_\tau | s_\tau)$$

$$= \mathbb{E}_{\tilde{Q}}[\log Q(s_\tau | \pi) - \log P(s_\tau | o_\tau, o_{t<\tau}, \pi) - \log P(o_\tau)]$$

We have expressed the generative model in terms of a prior over outcomes that does not depend upon the policy (last term).

$$\approx \mathbb{E}_{\tilde{Q}}[\log Q(s_\tau | \pi) - \log Q(s_\tau | o_\tau, \pi) - \log P(o_\tau)]$$

it is a mystery why $o_{t<\tau}$ disappeared

We shouldn't care about this approximation. It is only useful to relate to other constructs that people use. It is not implemented in the code.

Different types of approximation

of the posterior, to be able to compute free-energy.
This is for perception, i.e. within a timestep.

Best source: neuronal message passing using mean-field, Bethe and marginal approximations

Mean-field approximation

-> variational message passing

Although this paper only talks about belief propagation, it actually uses variational message passing and the simulations use marginal message passing. Variational message passing performs less well than marginal or belief propagation.

$$Q(\tilde{s} \mid \pi) := \prod_{\tau} Q(s_\tau \mid \pi)$$

Bethe approximation

-> Belief propagation

Better performance: gets the true posterior for non-cyclic generative models. For cyclic generative models it generalises *loopy belief propagation* and gives approximations. Less biologically plausible and gives silly answers for some kinds of generative models with loops.

$$Q(\tilde{s} \mid \pi) := \prod_{\tau} Q(s_\tau \mid \pi) \left(\prod_{\tau} \frac{Q(s_\tau, s_{\tau-1} \mid \pi)}{Q(s_\tau \mid \pi) Q(s_{\tau-1} \mid \pi)} \right)$$

Marginal approximation

-> Marginal message passing

Happy middle ground

Belief propagation

=Sum-product algorithm

Comprises Kalman filter, Bayesian filter and many AI algorithms as special case.

Updates (variational message passing)

Generative model

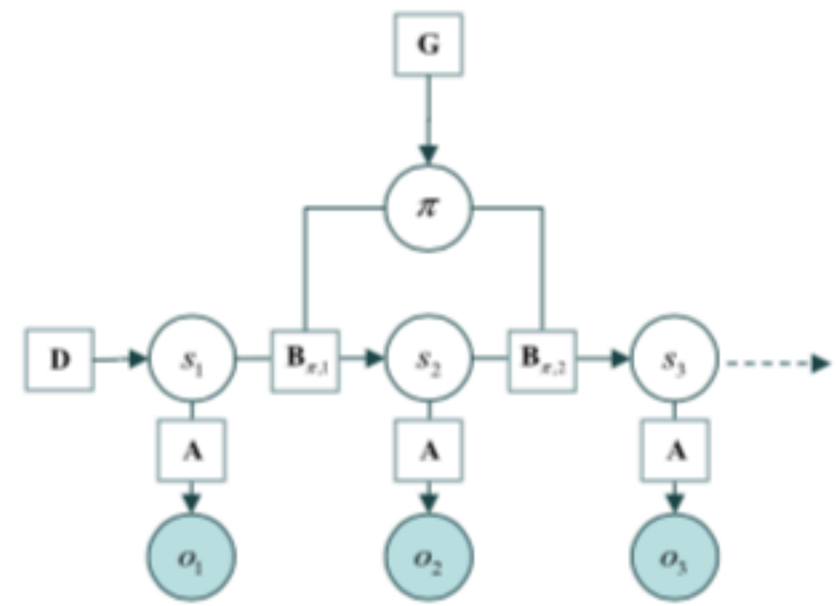
$$P(o_{1:\tau}, s_{1:\tau}, \pi) = P(s_1)P(\pi)\prod_\tau P(o_\tau \mid s_\tau)P(s_\tau \mid s_{\tau-1}, \pi)$$

Factors (likelihood and empirical priors)

1. $P(o_\tau \mid s_\tau) = Cat(\mathbf{A})$
2. $P(s_{\tau+1} \mid s_\tau, \pi) = Cat(\mathbf{B}_{\pi,\tau})$
3. $P(s_1) = Cat(\mathbf{D})$
4. $P(\pi) = \sigma(-\mathbf{G})$

Approximate posterior

$$Q(s_\tau \mid \pi) = Cat(\mathbf{s}_{\pi,\tau})$$
$$Q(\pi) = Cat(\boldsymbol{\pi})$$



Belief updating

$$\mathbf{s}_{\pi,\tau} = \sigma(\overset{2}{\ln \mathbf{B}_{\pi,\tau-1}\mathbf{s}_{\pi,\tau-1}} + \overset{3}{\ln \mathbf{B}_{\pi,\tau} \cdot \mathbf{s}_{\pi,\tau+1}} + \overset{4}{\ln \mathbf{A} \cdot o_\tau})$$

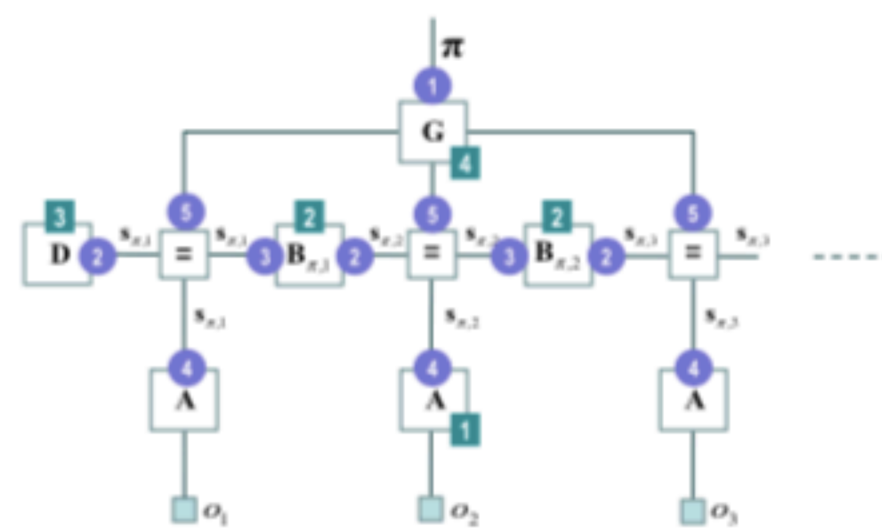
1. $\boldsymbol{\pi} = \sigma(-\mathbf{G})$

5. $\mathbf{G}_\pi = \sum_\tau \mathbf{o}_{\pi,\tau} \cdot (\ln \mathbf{o}_{\pi,\tau} + \mathbf{C}_\tau + \mathbf{H} \cdot \mathbf{s}_{\pi,\tau})$

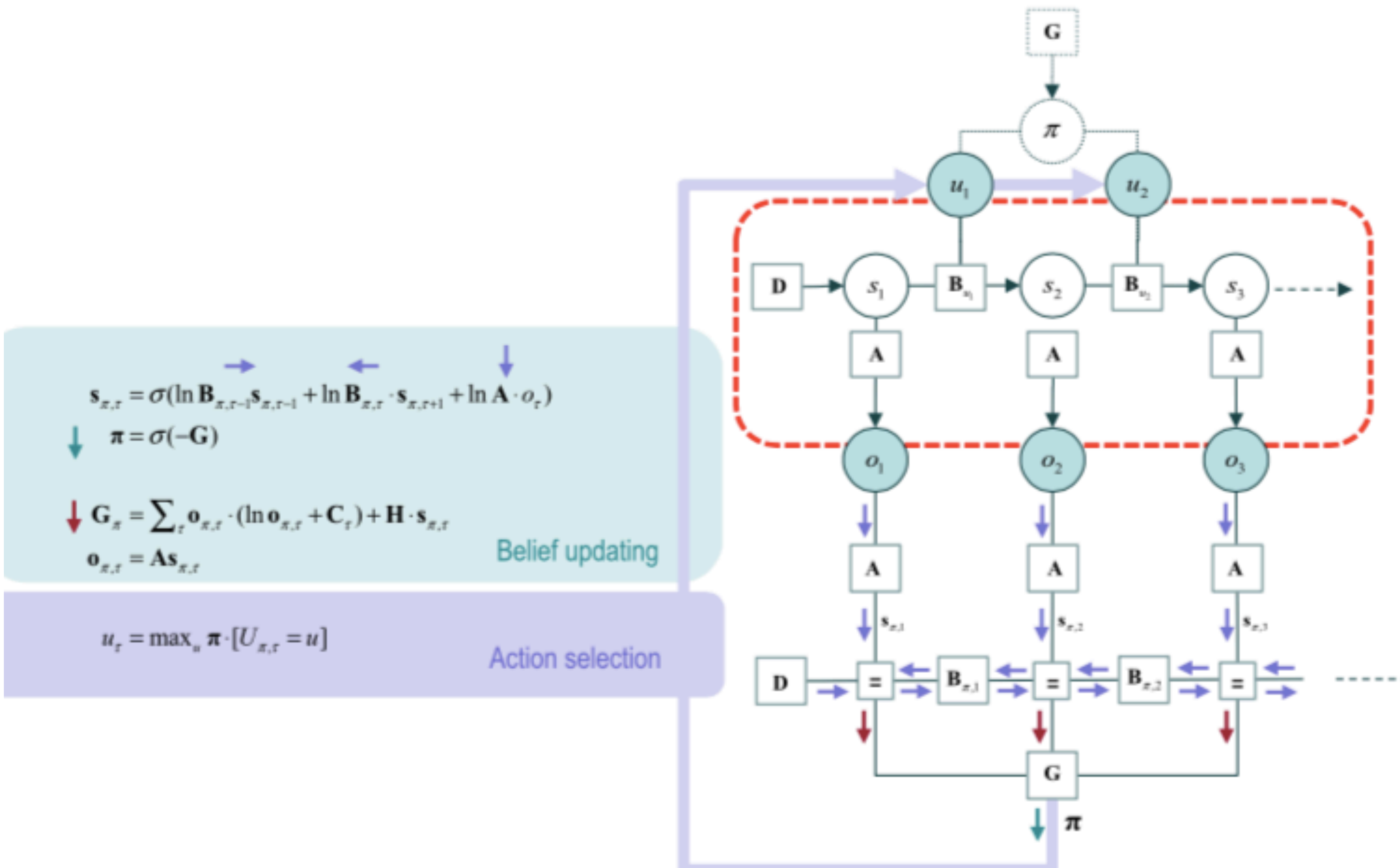
$\mathbf{o}_{\pi,\tau} = \mathbf{A}\mathbf{s}_{\pi,\tau}$

Action selection

$$u_t = \max_u \boldsymbol{\pi} \cdot [U_{\pi,t} = u]$$

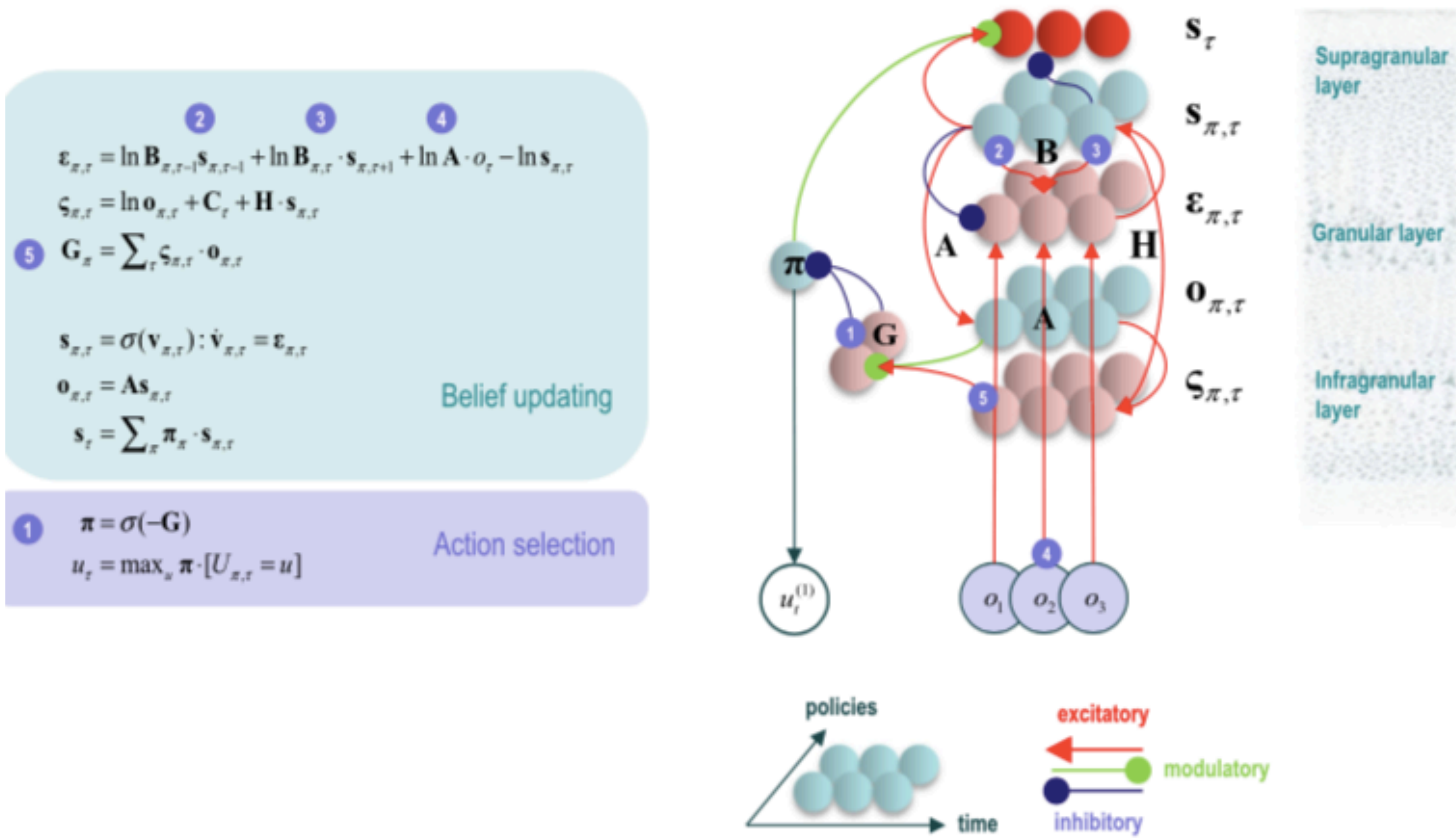


Generative model and generative process



Action selection: Selects the action corresponding to the most likely policy (= the one that minimises expected free-energy)

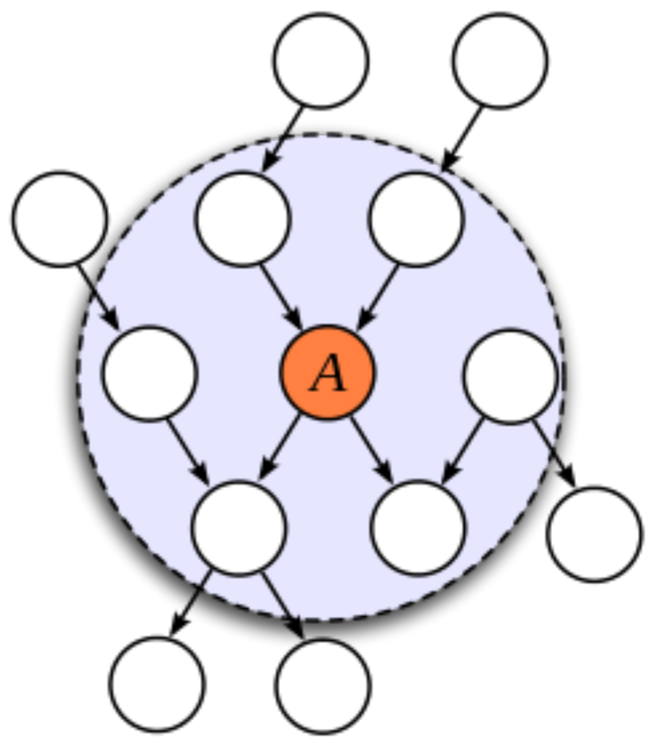
Neuronal implementation



Slightly speculative but based on empirical findings of neuronal structure and function. Can also lesion the model and see what pathological behaviour arises as a consequence to match it to corresponding areas in the brain.

Markov blanket

Bayesian network



$$P(A \mid \partial A, B) = P(A \mid \partial A)$$

Conjugate prior

(to the likelihood)

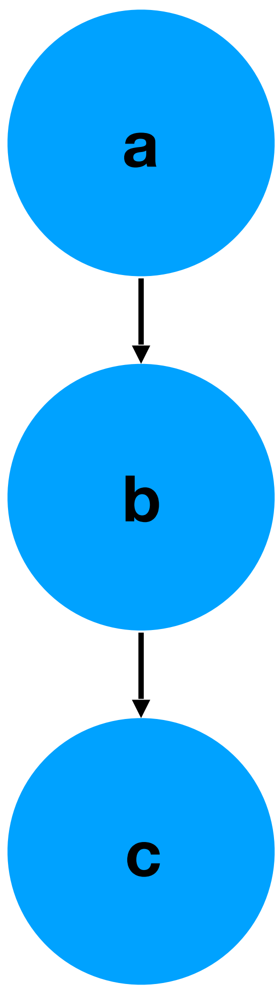
=If the posterior and prior are in the same family

$$p(x|y) \propto p(y|x)p(x)$$

e.g. Gaussian, gaussian

Categorical, Dirichlet

Deep hierarchical models



e.g.

$$a \sim \mathcal{N}(0,1)$$

$$b \,|\, a \sim \mathcal{N}(a, 1)$$

$$c \,|\, b, a \sim \mathcal{N}(b, 1)$$

$p(a)$ "Hyperprior"

$p(b \,|\, a)$ "Prior"

$p(c \,|\, b, a)$ "Likelihood"

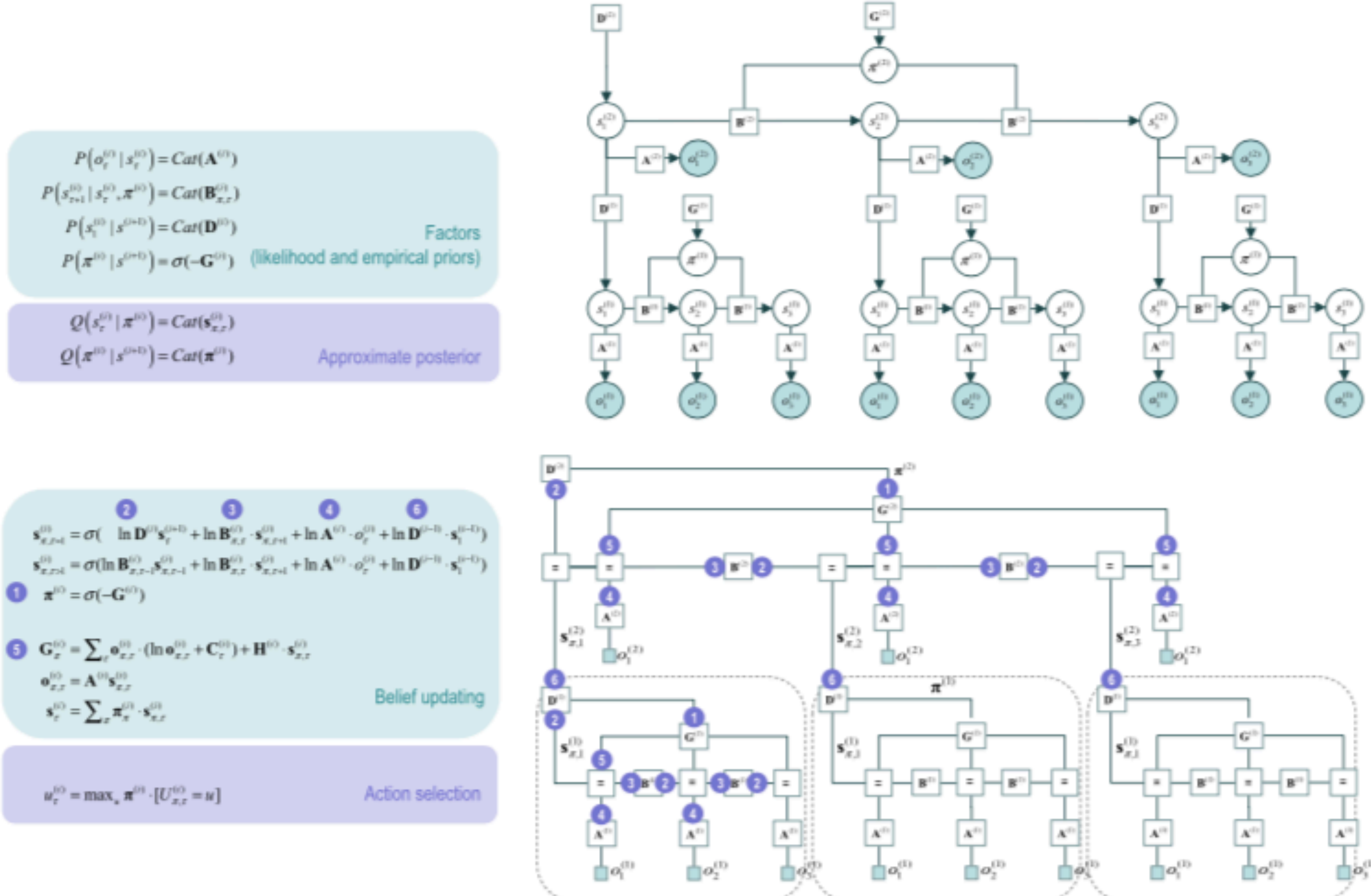
(more hyper priors if more layers)

Model inversion <=> Bayesian inference
i.e. inferring a,b from c

$$p(a, b \,|\, c) \propto p(c \,|\, b, a)p(b, a) = p(c \,|\, b)p(b, a) = p(c \,|\, b)p(b \,|\, a)p(a)$$

(more factors at the end if more layers, but generalises nicely)

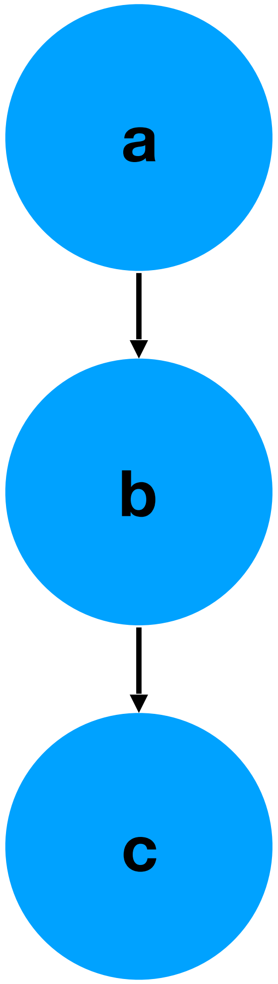
Deep hierarchical models



Higher level: slower timescales, lower level: faster timescales

Mixed models

A discrete model sitting on top of a continuous model



e.g.

$a \sim Dir(x)$

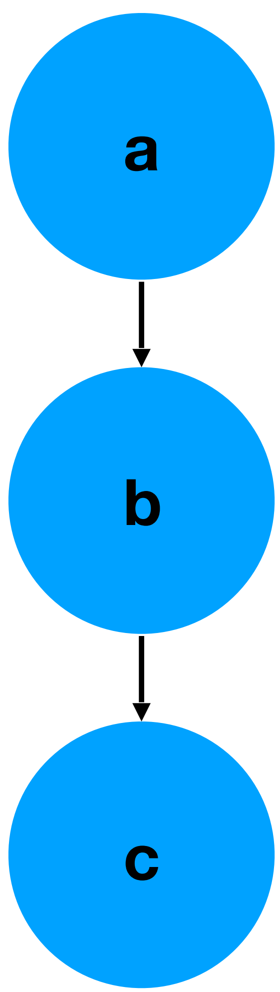
$b \sim Cat(a)$

$c \sim \mathcal{N}(\mu_b, 1)$

$$p(a, b \mid c) \propto p(c \mid b)p(b \mid a)p(a)$$

Continuous state-space models

Generalised coordinates of motion
~ coefficients of Taylor expansions of trajectories



e.g.

$$a \sim \mathcal{N}(0,1)$$

$$b \mid a \sim \mathcal{N}(a, 1)$$

$$c \mid b, a \sim \mathcal{N}(b, 1)$$