Active Inference: A Process Theory

Rafik Hadfi

May 18th, 2017

Summary

1. Introduction
2. Prerequisites
3. Active inference
4. Inference process
5. Simulation, T-maze foraging
6. Another usage

Introduction

- Active Inference, free energy minimization

Introduction

- Active Inference, free energy minimization

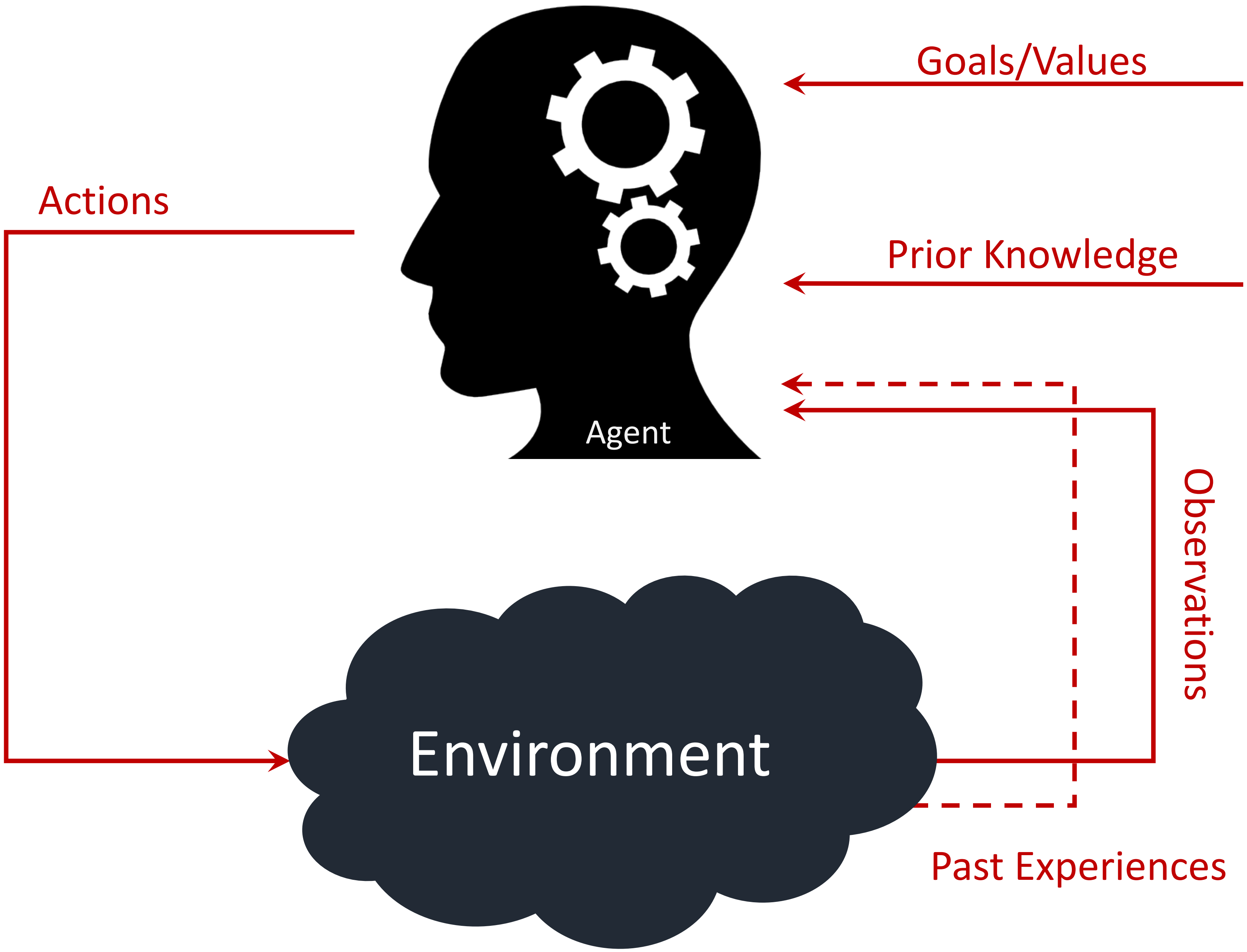
- Can it guide the behavior and evolution of an (artificial) agent?
- Is it a well-principled agent theory?

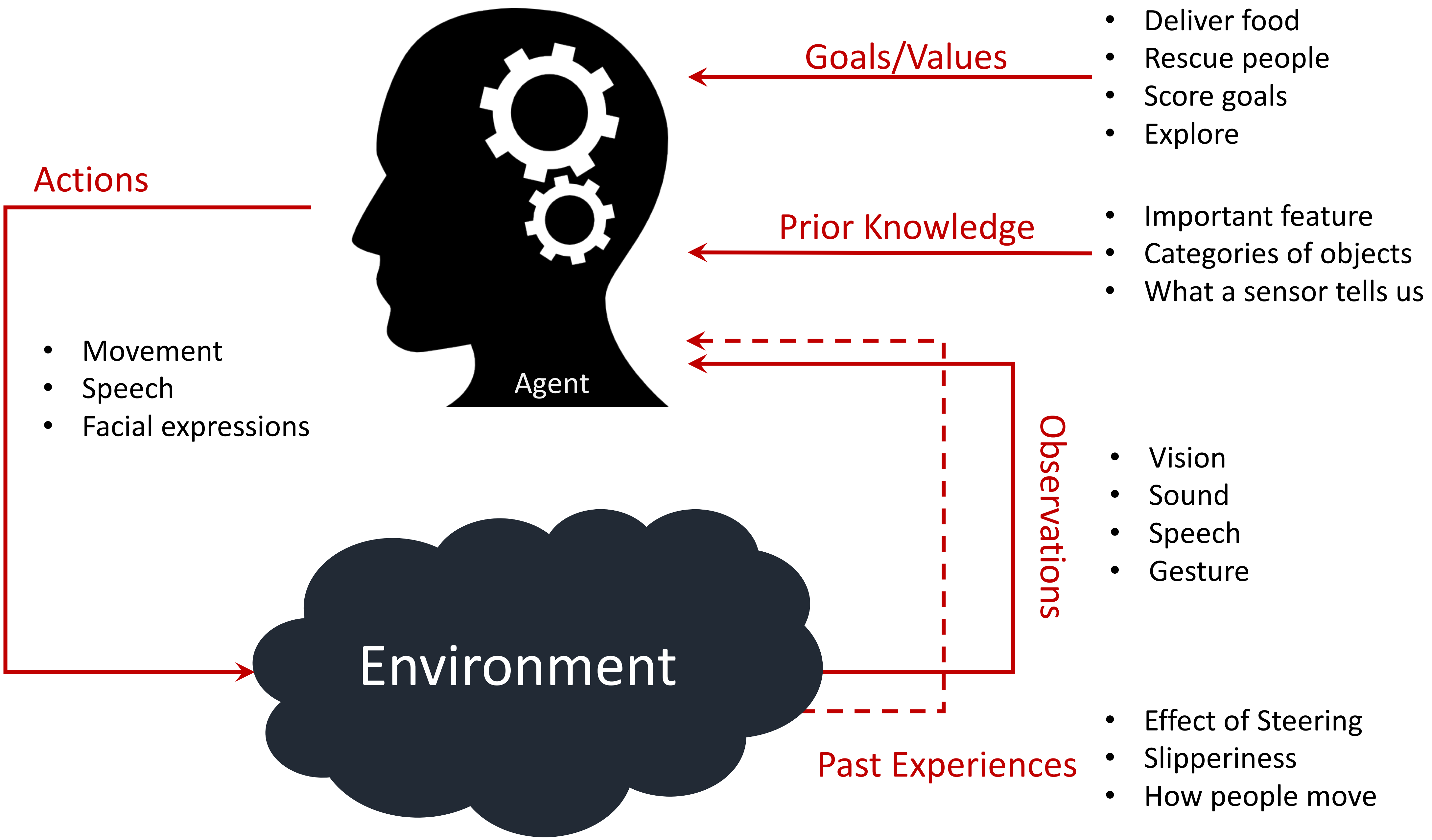
Introduction

- Active Inference, free energy minimization

- Can it guide the behavior and evolution of an (artificial) agent?
- Is it a well-principled agent theory?

- General problem of decision making or planning
  - When there is uncertainty about the outcomes, states, and observations
  - When the environment is dynamic


Goals/Values
Actions
Prior Knowledge
Agent
Observations
Environment
Past Experiences



Goals/Values
Deliver food
Rescue people
Score goals
Explore
Actions
Prior Knowledge
Important feature
Categories of objects
What a sensor tells us
Movement
Speech
Facial expressions
Agent
Observations
Vision
Sound
Speech
Gesture
Environment
Effect of Steering
Slipperiness
How people move
Past Experiences

Prerequisites

- Decision making in situations where outcomes are partly random and subject to uncertainty

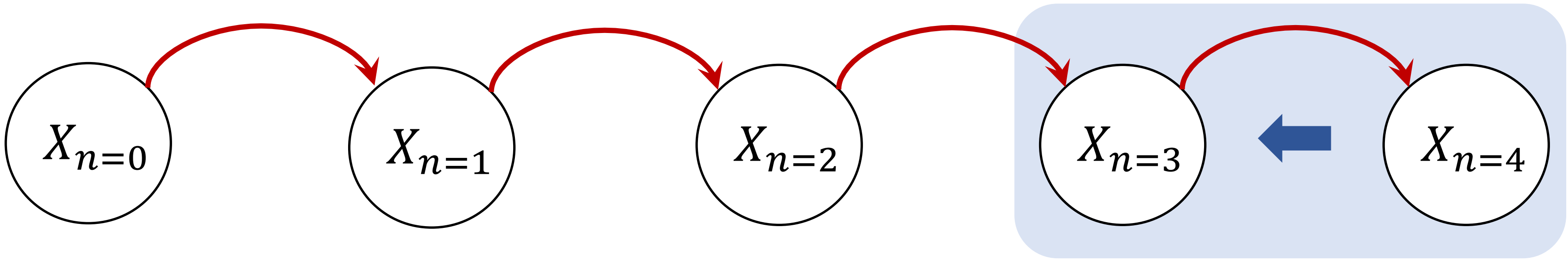
Prerequisites

- Decision making in situations where outcomes are partly random and subject to uncertainty
- Markov Process

Prerequisites

- Decision making in situations where outcomes are partly random and subject to uncertainty
- Markov Process
  - is a stochastic process that satisfies the Markov property (memorylessness), where one can make predictions for the future of the process based solely on its present state

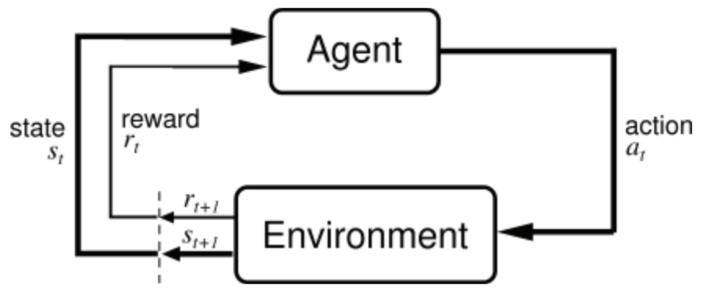
$$P(X_n = x_n | X_{n-1} = x_{n-1}, X_{n-2} = x_{n-2}, \dots, X_0 = x_0) = P(X_n = x_n | X_{n-1} = x_{n-1})$$

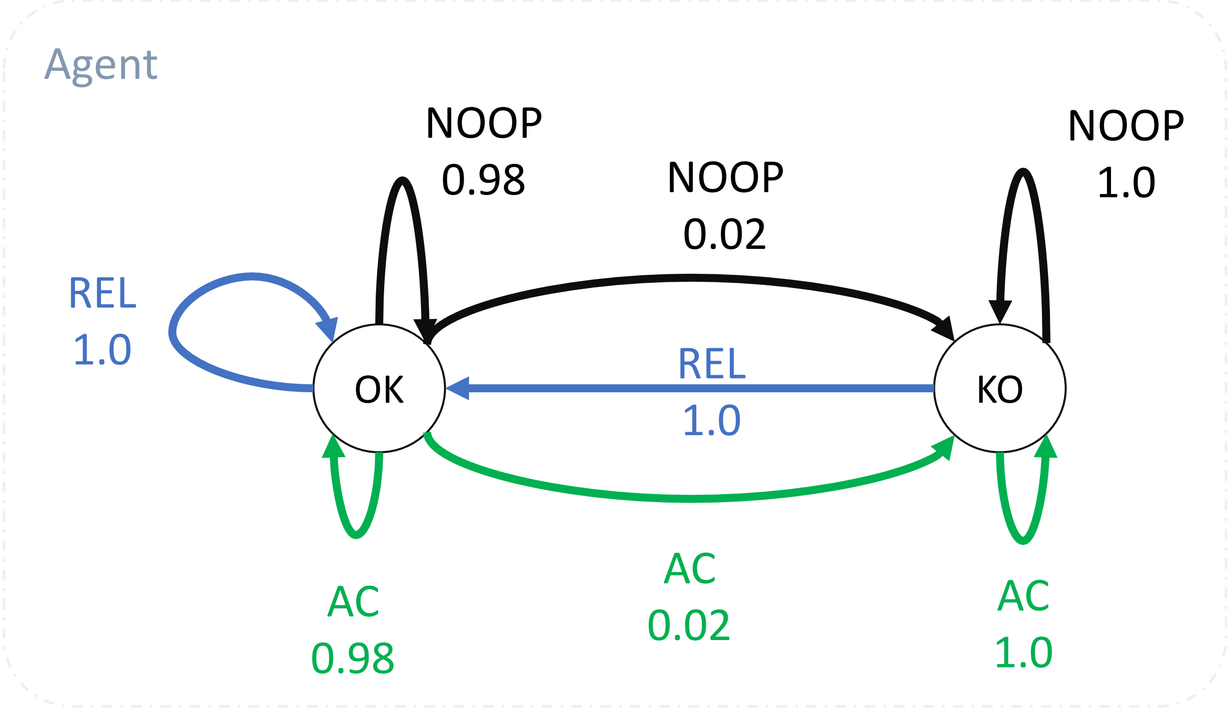


Prerequisites

MDP
$(S, A, \mathbb{P}, R)$

- Set of states $S = \{OK, KO\}$
- Set of Actions $A = \{NOOP, AC, REL\}$
- Reward function $R_a(s, s'): A \times S \times S \rightarrow \mathbb{R}$
- State transitions $\mathbb{P}(s' | a, s)$

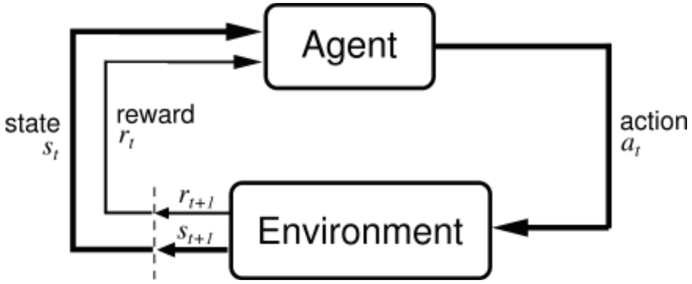


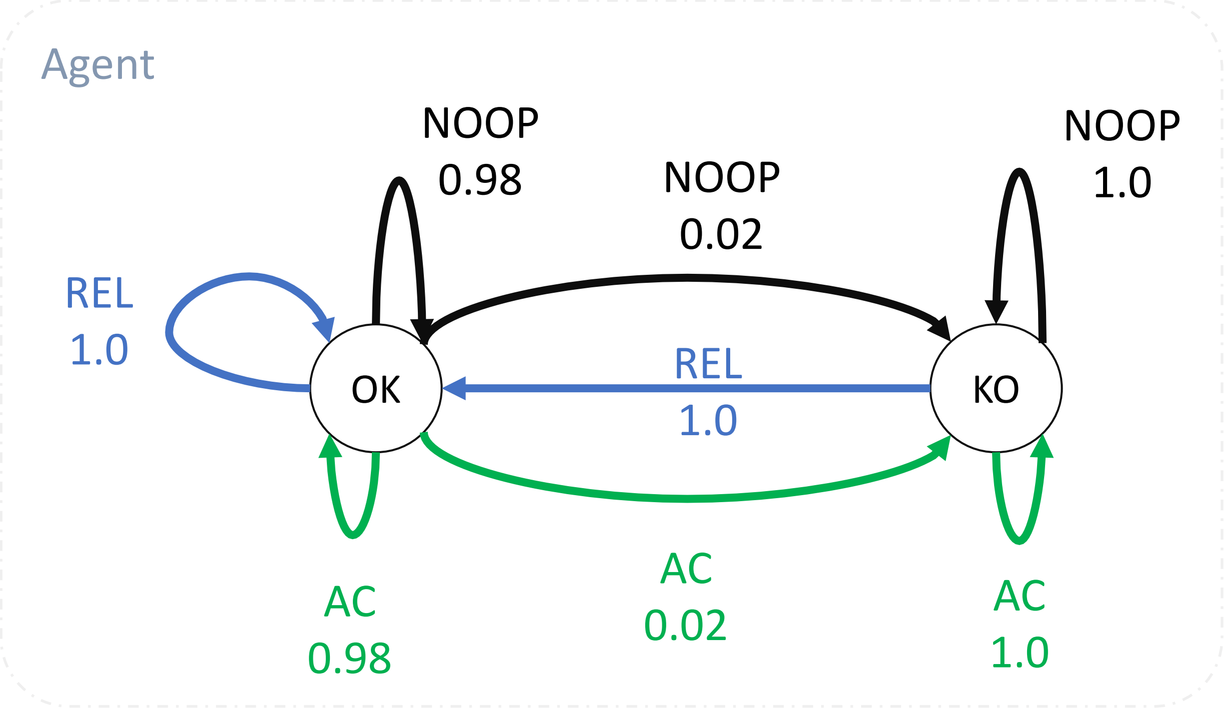


Prerequisites

MDP
$(S, A, \mathbb{P}, R)$

- Set of states $S = \{OK, KO\}$
- Set of Actions $A = \{NOOP, AC, REL\}$
- Reward function $R_a(s, s'): A \times S \times S \rightarrow \mathbb{R}$
- State transitions $\mathbb{P}(s' | a, s)$



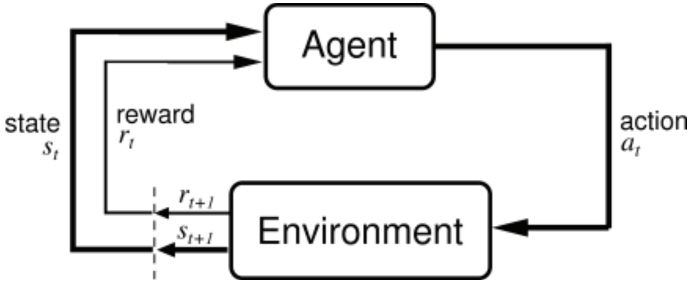


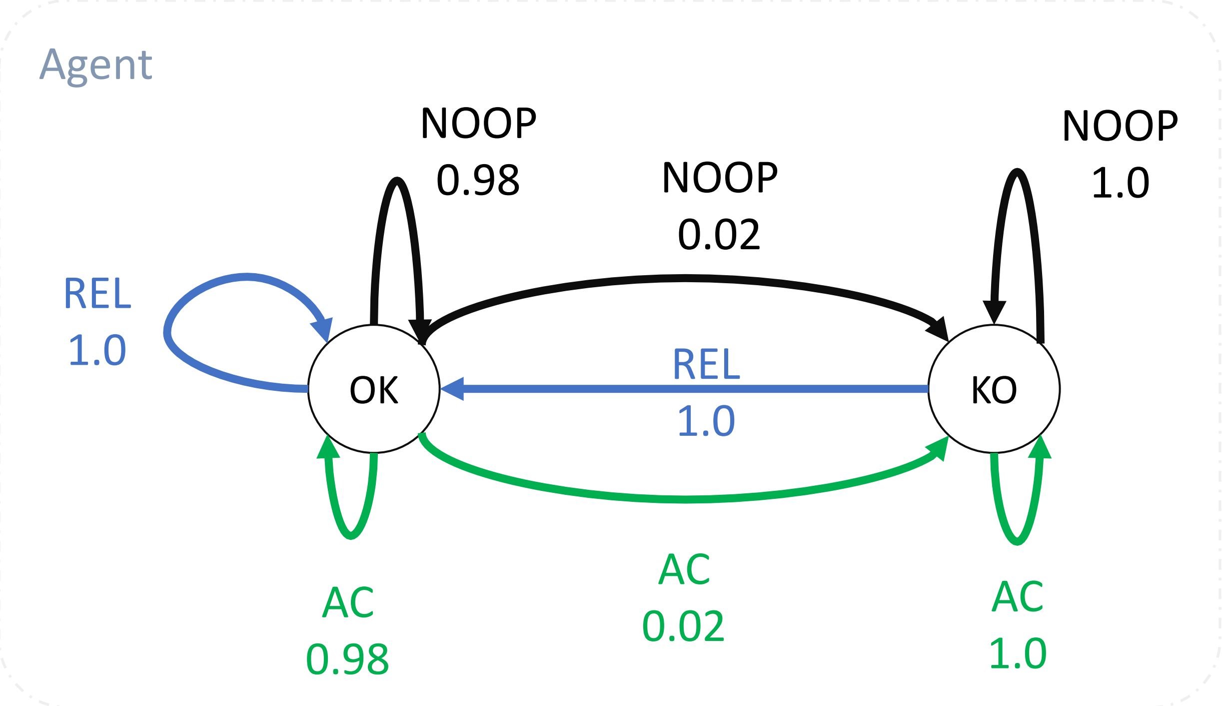
Question: What will be the best decision strategy in the long term: If the agent is KO, are we better off restarting or relocating?

Prerequisites

MDP
$(S, A, \mathbb{P}, R)$

- Set of states $S = \{OK, KO\}$
- Set of Actions $A = \{NOOP, AC, REL\}$
- Reward function $R_a(s, s'): A \times S \times S \rightarrow \mathbb{R}$
- State transitions $\mathbb{P}(s' | a, s)$





Question: What will be the best decision strategy in the long term: If the agent is KO, are we better off restarting or relocating?

Solution: Optimal policy $\pi: S \rightarrow A$

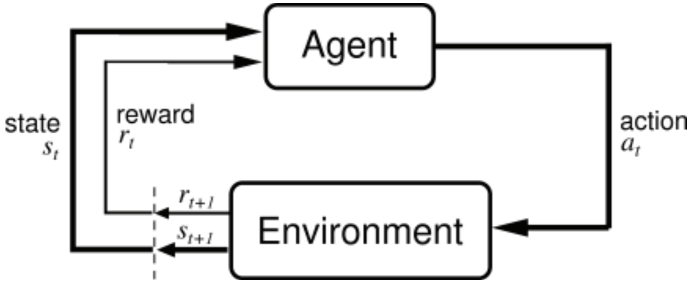
π maximizes cumulative reward $\sum_{t=0}^{\infty} \gamma^t R_{a_t}(s_t, s_{t+1})$, with $a_t = \pi(s_t)$, using a recursive algorithm:

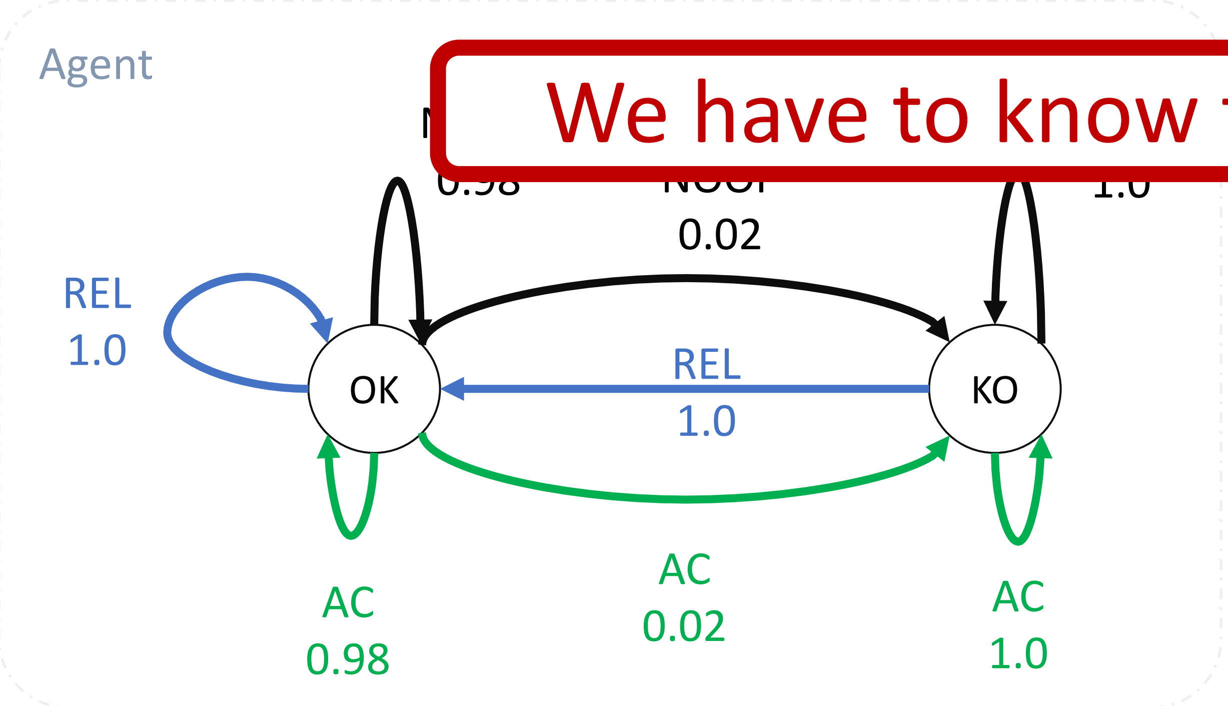
$$\pi(s) := argmax_a \sum_{s'} \mathbb{P}(s' | a, s)[R_a(s, s') + \gamma V(s')]$$
$$V(s) := \sum_{s'} \mathbb{P}(s' | \pi(s), s)[R_{\pi(s)}(s, s') + \gamma V(s')]$$

Prerequisites

MDP $(S, A, \mathbb{P}, R)$

- Set of states $S = \{OK, KO\}$
- Set of Actions $A = \{NOOP, AC, REL\}$
- Reward function $R_a(s, s'): A \times S \times S \rightarrow \mathbb{R}$
- State transitions $\mathbb{P}(s' \mid a, s)$





We have to know the agent states!

… decision strategy in … , are we better off … restarting or relocating?

Solution: Optimal policy $\pi: S \rightarrow A$

π maximizes cumulative reward $\sum_{t=0}^{\infty} \gamma^t R_{a_t}(s_t, s_{t+1})$, with $a_t = \pi(s_t)$, using a recursive algorithm

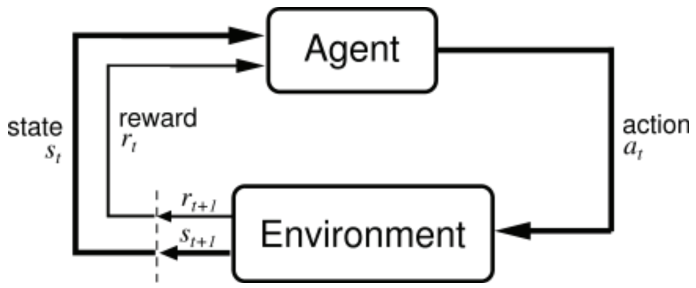
$$\pi(s) := argmax_a \sum_{s'} \mathbb{P}(s' | a, s)[R_a(s, s') + \gamma V(s')]$$
$$V(s) := \sum_{s'} \mathbb{P}(s' | \pi(s), s)[R_{\pi(s)}(s, s') + \gamma V(s')]$$

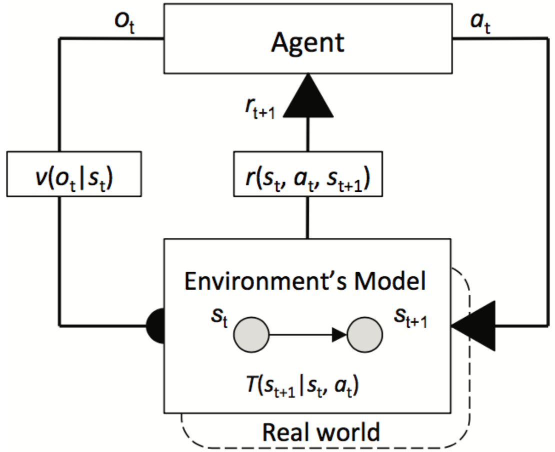
Prerequisites

Observability of the states

	Full Observability	Partial Observability
No Actions	Markov Process	HMM
Actions	MDP	POMDP

Do we have control over the state transitions?



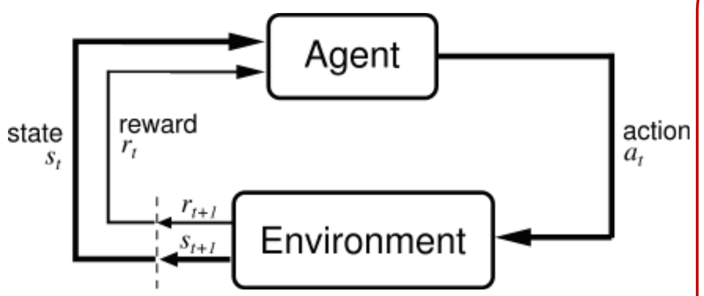


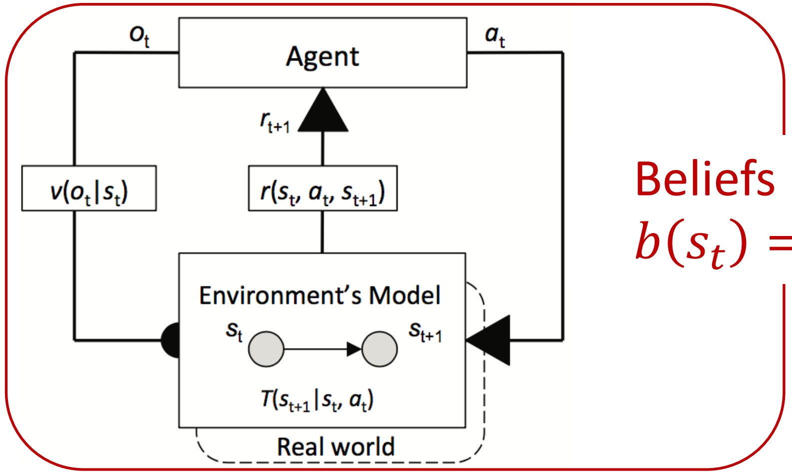
Prerequisites

Observability of the states

	Full Observability	Partial Observability
No Actions	Markov Process	HMM
Actions	MDP	POMDP

Do we have control over the state transitions?

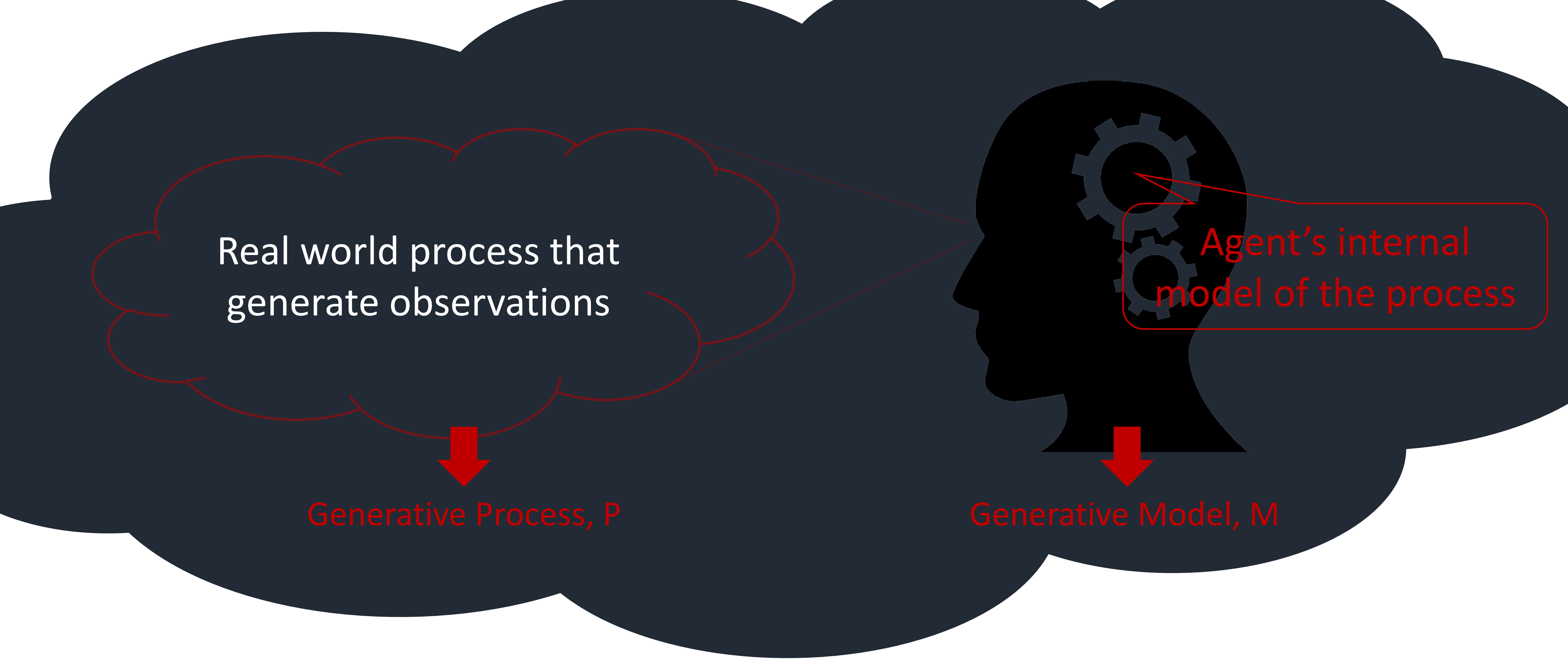




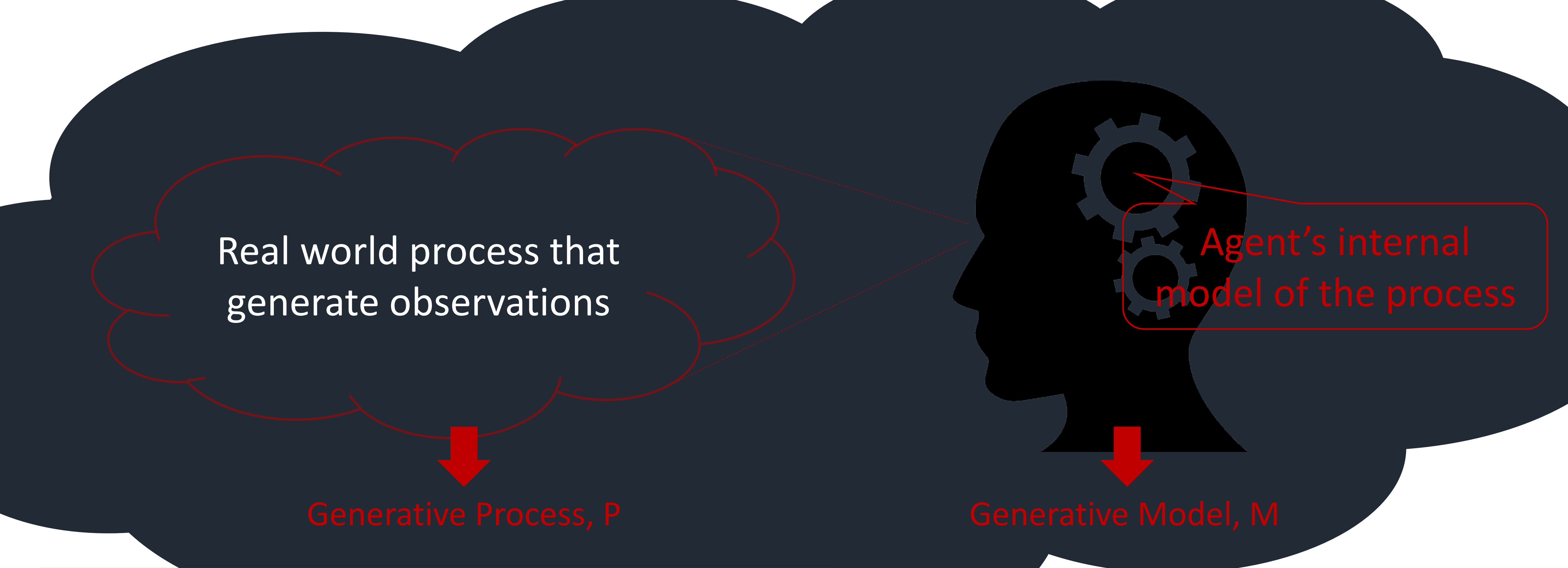
Beliefs update

$b(s_t) = \mathbb{P}(s_t)$

Active Inference



Sensory observations generated by P are observed by the agent while the agent is acting on the world to change P



P describes transitions among states in the world that generate observed outcomes. These states are referred to as hidden because they cannot be observed directly. Their transitions depend on action, which depends on posterior beliefs about the next state. These beliefs are formed using a generative model of how observations are generated.

M describes what the agent believes about the world, where beliefs about hidden states and policies are encoded by expectations.

Real world process that generate observations

Generative process $R(\tilde{o}, \tilde{s}, \tilde{u}, \eta)$ that generates probabilistic outcomes $o \in O$ from (hidden) states $s \in S$ and actions $a \in \Upsilon$

The generative model $P(\tilde{o}, \tilde{s}, \tilde{u}, \eta)$ can be parametrized in a general way using $\eta = \{\mathrm{a}, \mathrm{b}, \mathrm{c}, \mathrm{d}, \beta\}$ over outcomes, states, and policies $\pi \in T$ where $\pi \in \{0, \ldots, K\}$ returns a sequence of actions $u_t = \pi(t)$

$$P(\tilde{o}, \tilde{s}, \tilde{u}, \eta) = P(\pi)P(\eta)\prod_{t=1}^{T} P(o_t|s_t)P(s_t|s_{t-1}, \pi)$$

$$P(o_t|s_t) = Cat(A)$$

$$P(s_t|s_{t-1}, \pi) = \mathrm{Cat}\left(\mathrm{B}(\mathrm{u} = \pi(t))\right)$$

$$P(s_1|s_0) = \mathrm{Cat}(D)$$

$$P(\pi) = \sigma\left(-\gamma . G(\pi)\right)$$

$$P(A) = Dir(\alpha)$$

$$P(B) = Dir(b)$$

$$P(D) = Dir(d)$$

$$P(\gamma) = \Gamma(1, \beta)$$

An approximate posterior over hidden states and parameters $x = (\tilde{s}, \pi, \eta)$ is expressed as

$$Q(x) = Q(s_1|\pi) \ldots Q(s_T|\pi)Q(\pi)Q(\boldsymbol{A})Q(\boldsymbol{B})Q(\boldsymbol{D})Q(\boldsymbol{\gamma})$$

P describes transitions among states in the world that generate observed outcomes. These states are referred to as hidden because they cannot be observed directly. Their transitions depend on action, which depends on posterior beliefs about the next state. These beliefs are formed using a generative model of how observations are generated.

M describes what the agent believes about the world, where beliefs about hidden states and policies are encoded by expectations.

Active Inference

$$(O, P, Q, R, S, T, \Upsilon)$$

- Finite set of outcomes, O
- Generative model $P(\tilde{o}, \tilde{s}, \tilde{u})$with parameters η over outcomes, states, and policies $\pi \in T$ where $\pi \in \{0, \ldots, K\}$ returns a sequence of actions $u_t = \pi(t)$
- An approximate posterior $Q(\tilde{s}, \pi, \eta) = Q(s_0|\pi) \ldots Q(s_T|\pi)\ Q(\pi)\ Q(\eta)$over states, policies and parameters with expectations $(s_0^\pi, \ldots, s_0^\pi, \pi, \eta)$
- Generative process $R(\tilde{o}, \tilde{s}, \tilde{u})$ that generates probabilistic outcomes $o \in O$ from hidden states $s \in S$ and actions $a \in \Upsilon$
- Finite set S of hidden states
- Finite set T of time-sensitive policies
- Finite set Υ of control states, or actions

Inference process

Agent

World

Inference process

Generative model $P(\tilde{o}, \tilde{\mathbf{s}}, \tilde{a}, \tilde{u}|m)$ connecting observations to hidden states

- Predictive model over observations
- Encodes optimal policies in term of prior beliefs about control states

Posterior beliefs $Q(\tilde{s}, \tilde{u}|\mu)$ about those states, parametrized by expectations

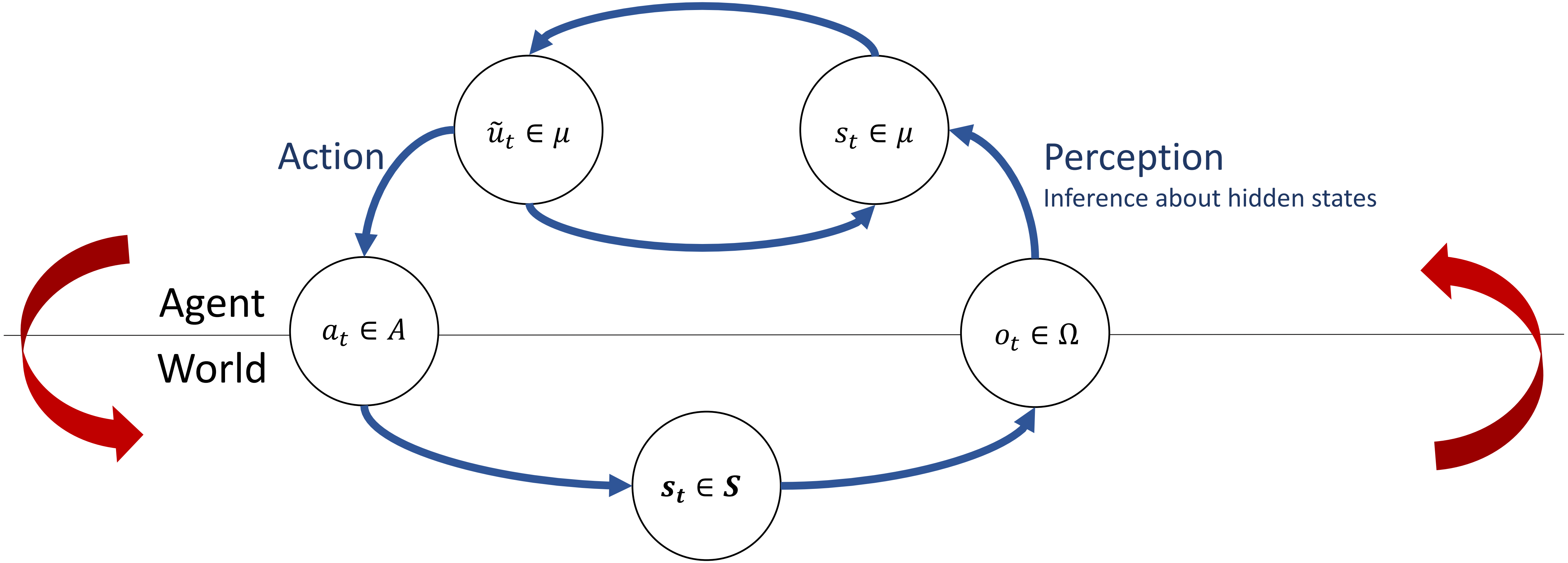
Agent

World

Distribution $R(\tilde{o}, \tilde{\mathbf{s}}, \tilde{a})$ over observations

- True states $\tilde{\mathbf{s}}$ control environmental responses but are never observed directly. Instead, the agent infer hidden states based on observations

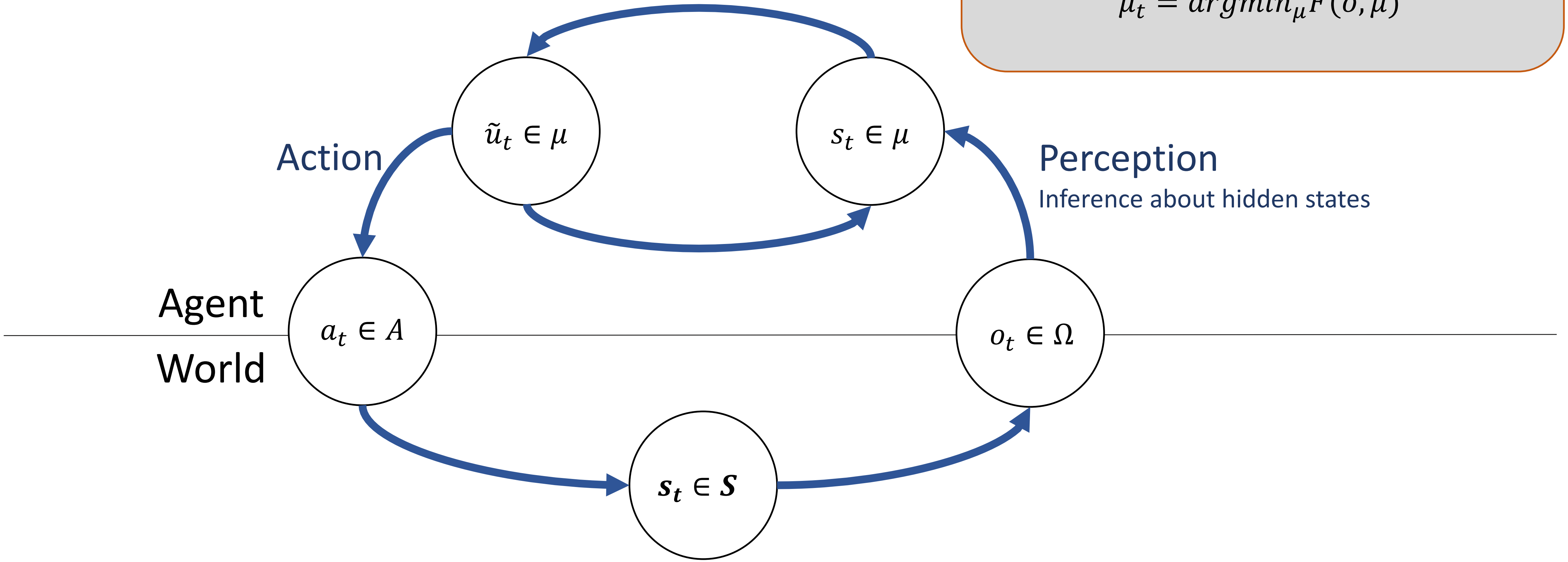
Inference process



Inference process

1 Find which states are most likely by optimizing expectation with regard to the free energy of observations

$$\mu_t = argmin_{\mu} F(\tilde{o}, \mu)$$

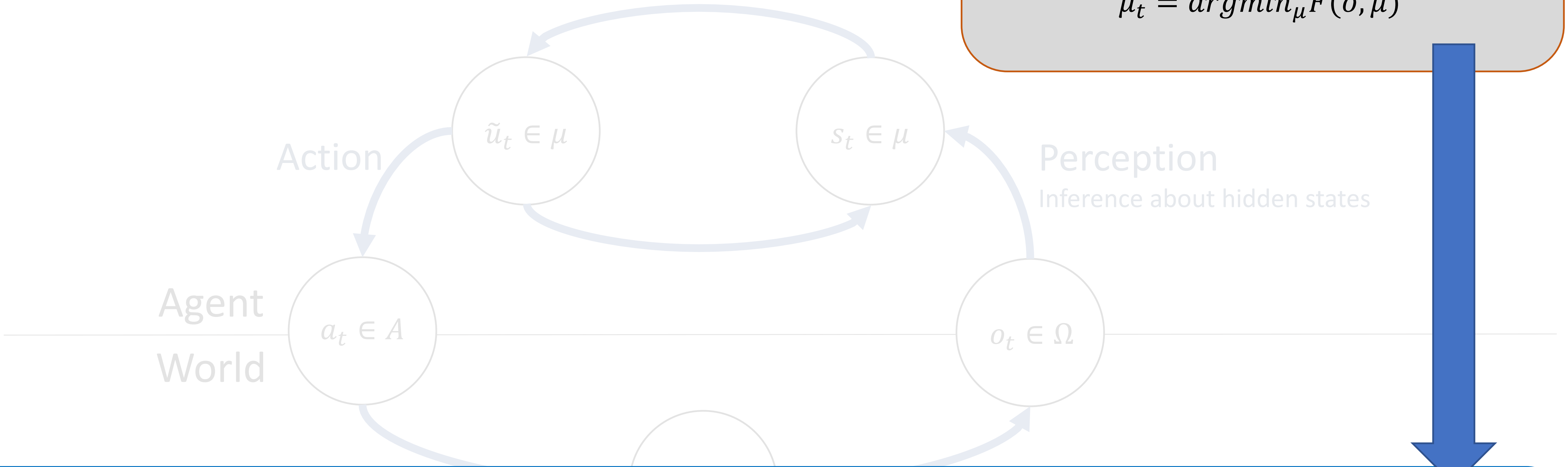


Inference process

Find which states are most likely by optimizing expectation with regard to the free energy of observations

1

$$\mu_t = argmin_\mu F(\tilde{o}, \mu)$$



Gibbs energy (expected under the approximate prior) - Entropy of the approximate prior

The reason why we call it free energy!

$$F(\tilde{o}, \mu) = \mathbb{E}_Q[-\ln P(\tilde{o}, \tilde{s}, \tilde{u}|m)] - H(P(\tilde{s}, \tilde{u}|\mu))$$

Inference process

1 Find which states are most likely by optimizing expectation with regard to the free energy of observations

$$\mu_t = argmin_\mu F(\tilde{o}, \mu)$$

Complexity - Accuracy

Minimizing free energy is the same as maximizing the expected log likelihood of observations or accuracy, while minimizing the divergence between the approximate posterior and prior beliefs about hidden variables. This divergence is known as model complexity.

$$F(\tilde{o}, \mu) \; = D_{KL}[Q(\tilde{s}, \tilde{u}|\mu)||P(\tilde{s}, \tilde{u}|\tilde{o})] + \mathbb{E}_Q[-\ln P(\tilde{o}|\tilde{s}, \tilde{u})]$$

Expected entropy of observations

Gibbs energy (expected under the approximate prior) - Entropy of the approximate prior

The reason why we call it free energy!

$$F(\tilde{o}, \mu) \; = \mathbb{E}_Q[-\ln P(\tilde{o}, \tilde{s}, \tilde{u}|m)] - H(P(\tilde{s}, \tilde{u}|\mu))$$

(Divergence + Surprise) or (Relative Entropy - log evidence)

Free energy is an upper bound on surprise, because $D_{KL}(.||.) \geq 0$ (Gibbs inequality)

$$F(\tilde{o}, \mu) = D_{KL}[Q(\tilde{s}, \tilde{u}|\mu)||P(\tilde{s}, \tilde{u}|\tilde{o})] - \ln \mathrm{P}(\tilde{o}|m)$$

Posterior (predictive) distribution over hidden states.

Prior (preferred) distribution over future outcomes.

Minimizing free energy corresponds to minimizing divergence between the approximate and true posterior.

Find which states are most likely by optimizing expectation with regard to the free energy of observations

$$\mu_t = argmin_{\mu} F(\tilde{o}, \mu)$$

Complexity - Accuracy

Minimizing free energy is the same as maximizing the expected log likelihood of observations or accuracy, while minimizing the divergence between the approximate posterior and prior beliefs about hidden variables. This divergence is known as model complexity.

$$F(\tilde{o}, \mu) = D_{KL}[Q(\tilde{s}, \tilde{u}|\mu)||P(\tilde{s}, \tilde{u}|\tilde{o})] + \mathbb{E}_Q[-\ln P(\tilde{o}|\tilde{s}, \tilde{u})]$$

Expected entropy of observations

Gibbs energy (expected under the approximate prior) - Entropy of the approximate prior

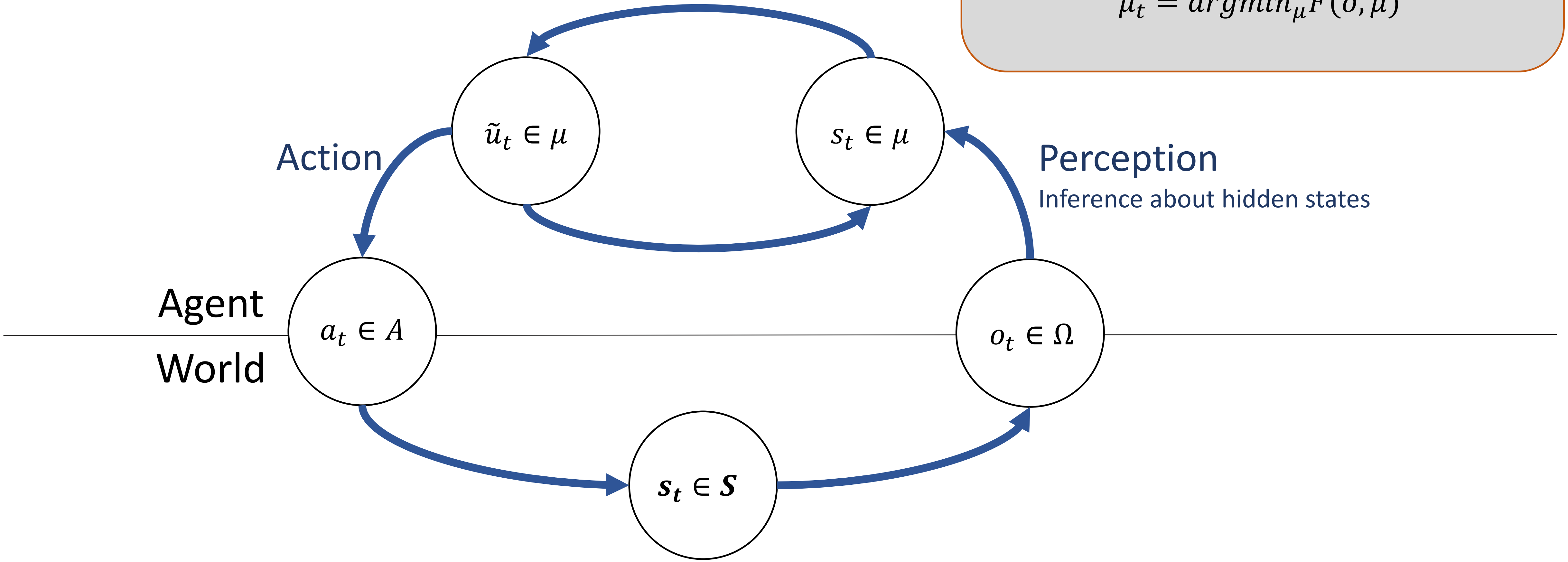
The reason why we call it free energy!

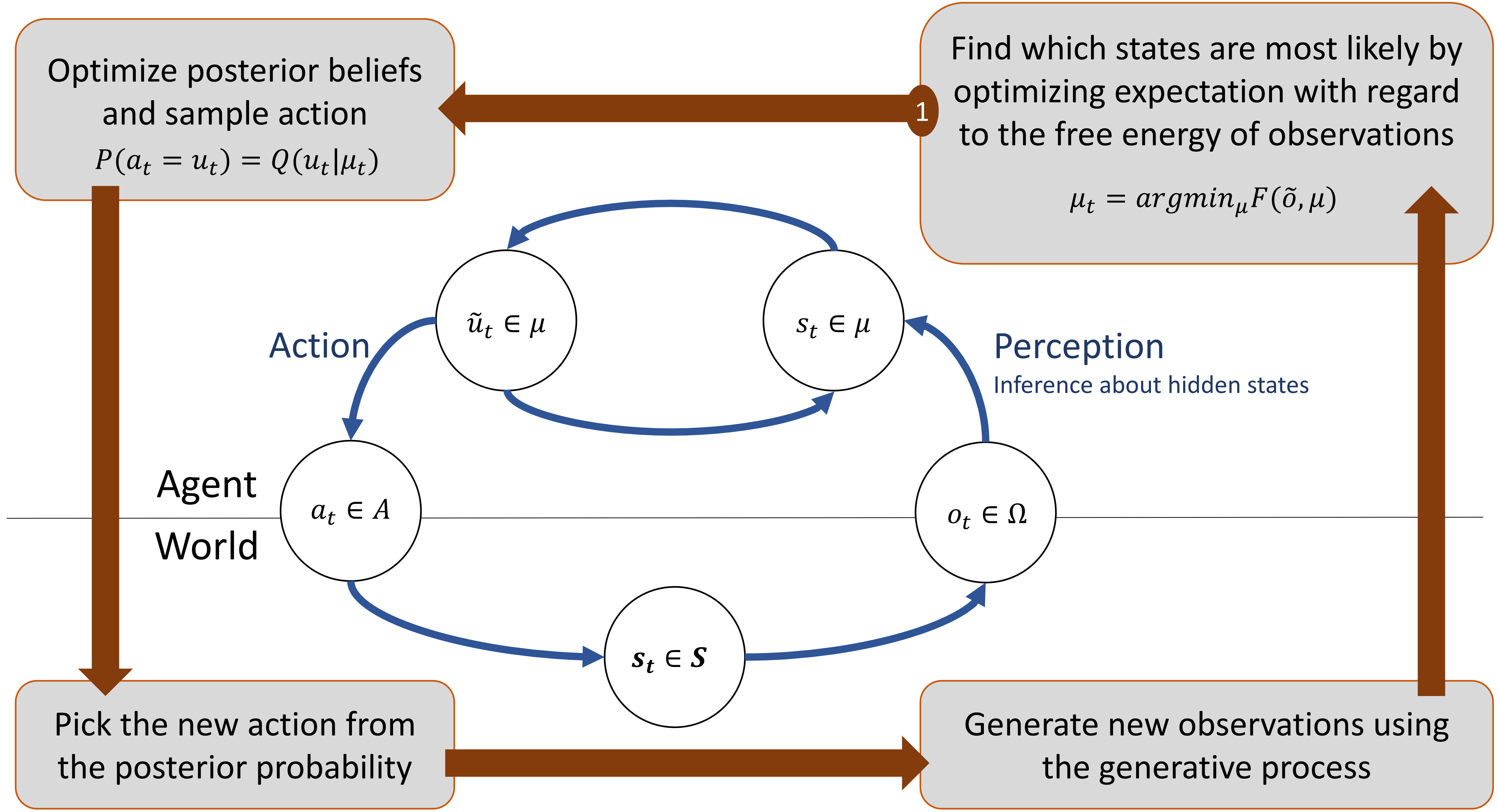
$$F(\tilde{o}, \mu) = \mathbb{E}_Q[-\ln P(\tilde{o}, \tilde{s}, \tilde{u}|m)] - H(P(\tilde{s}, \tilde{u}|\mu))$$

Inference process

1 Find which states are most likely by optimizing expectation with regard to the free energy of observations

$$\mu_t = argmin_\mu F(\tilde{o}, \mu)$$





Optimize posterior beliefs and sample action
$P(a_t = u_t) = Q(u_t|\mu_t)$
1
Find which states are most likely by optimizing expectation with regard to the free energy of observations
$\mu_t = argmin_\mu F(\tilde{o}, \mu)$
$\tilde{u}_t \in \mu$
$s_t \in \mu$
Action
Perception
Inference about hidden states
Agent
World
$a_t \in A$
$o_t \in \Omega$
$\boldsymbol{s_t} \in \boldsymbol{S}$
Pick the new action from the posterior probability
Generate new observations using the generative process

Optimize posterior beliefs and sample action

$$P(a_t = u_t) = Q(u_t|\mu_t)$$

1

Find which states are most likely by optimizing expectation with regard to the free energy of observations

$$\mu_t = argmin_\mu F(\tilde{o}, \mu)$$

When expressed using a policy

$$(\tilde{s}^*, \tilde{\pi}^*) = argmin\, F(\tilde{o}, \tilde{s}, \tilde{\pi})$$

$$P(a_t = u_t) = Q(u_t|\tilde{\pi}^*)$$

the negative free energy of the approximate posterior predictive density becomes

$$Q_\tau(\pi) = \mathbb{E}_{Q(o_\tau, s_\tau|\pi)}[\ln P(o_\tau, s_\tau|\pi)] + H(Q(s_\tau|\pi))$$

A policy is a priori more likely if it has high quality or if its expected free energy is small.

→ Heuristically, the agent believes they will pursue policies that minimize the expected free energy of outcomes and implicitly minimize their surprise about those outcomes.

Optimize posterior beliefs and sample action

$P(a_t = u_t) = Q(u_t|\mu_t)$

1

Find which states are most likely by optimizing expectation with regard to the free energy of observations

$\mu_t = argmin_\mu F(\tilde{o}, \mu)$

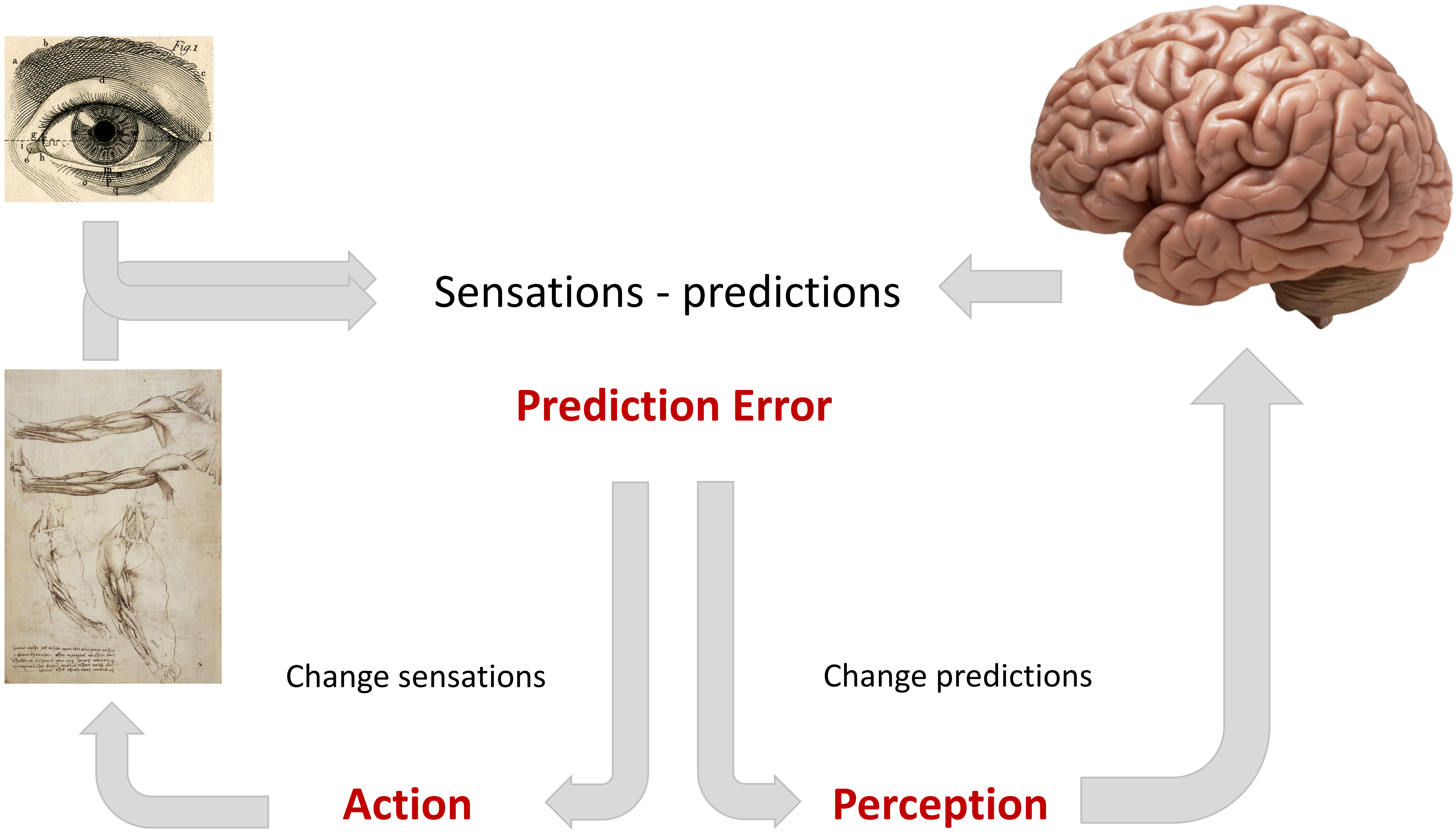
Under the generative model of the future, the quality of a policy, $Q_\tau(\pi)$, can be rewritten as

$$Q_\tau(\pi) \ = \mathbb{E}_{Q(o_\tau|\pi)}[\ln P(o_\tau|m)] + \mathbb{E}_{Q(o_\tau|\pi)}[D_{KL}(Q(s_\tau|o_\tau,\pi)||Q(s_\tau||\pi))]$$

Extrinsic value   Epistemic value

Extrinsic value is the utility $\mathrm{C}(o_\tau|m) = \ln P(o_\tau|m)$ of an outcome expected under the posterior predictive distribution. It encodes the preferred outcomes that give the goal-directed behavior.

Epistemic (intrinsic) value is the expected information gain under predicted outcomes. It reports the reduction in uncertainty about hidden states afforded by observations. The information gain is smallest when the posterior predictive distribution is not informed by new observations.

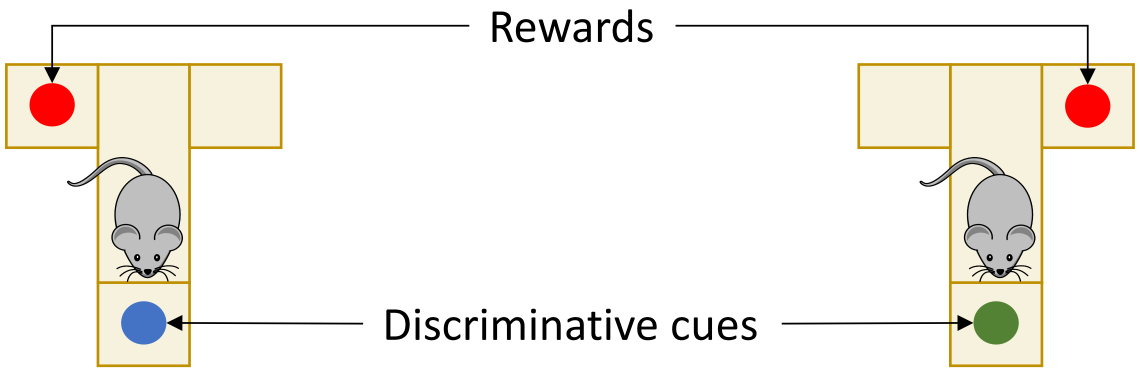

Sensations - predictions
Prediction Error
Change sensations
Change predictions
Action
Perception

Simulation

- Simulating foraging in a T-maze using a hierarchical generative model?

Simulation

- Simulating foraging in a T-maze using a hierarchical generative model?
  - The agent starts in the center, where either the right or left arms are baited with a reward.
  - The lower arm contains a discriminative cue that tells the agent whether the reward is in the upper right or left arm.
  - The agent can make only two moves.
  - The agent cannot leave the baited arms after they are entered.
  - The optimal behavior is to first go to the lower arm to find where the reward is located and then retrieve the reward at the cued location.

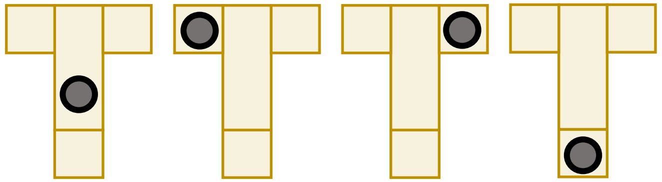


Simulation

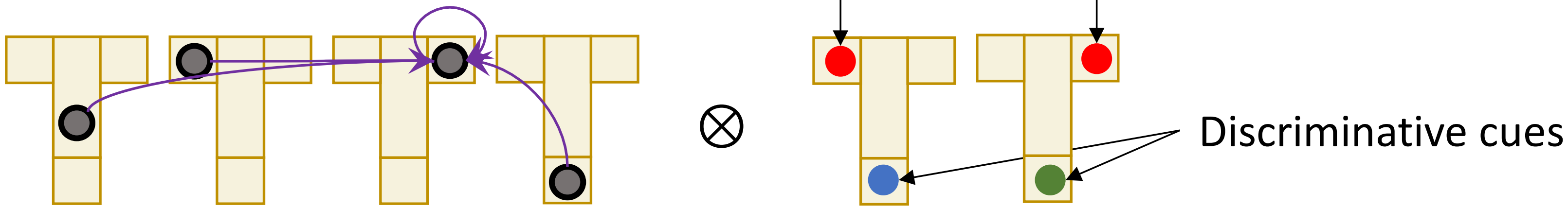
- Simulating foraging in a T-maze using a hierarchical generative model?
  - Translate into a POMDP

Simulation

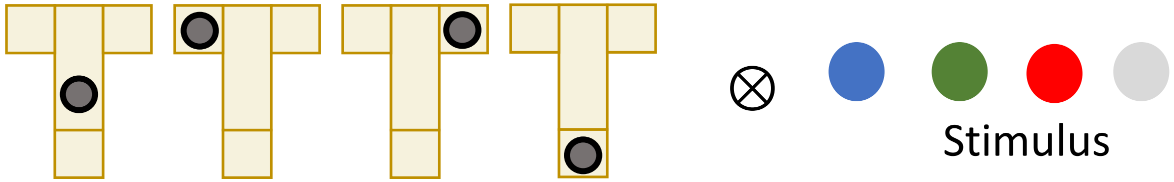
- Control states
  $u \in U$



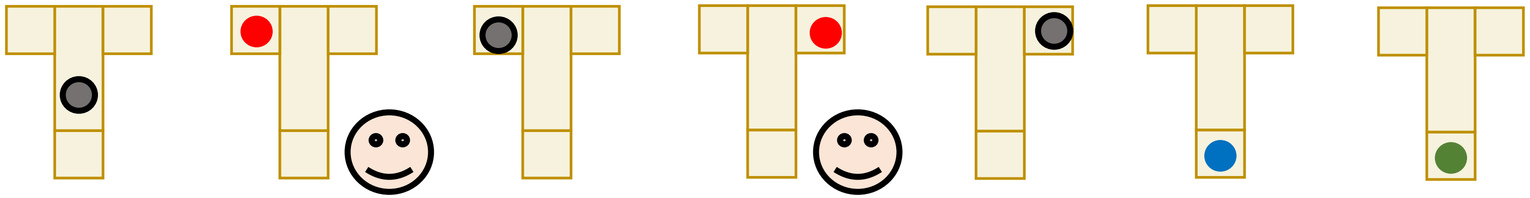
- Hidden states
  $s = s_l \otimes s_c \in S$



- Observations

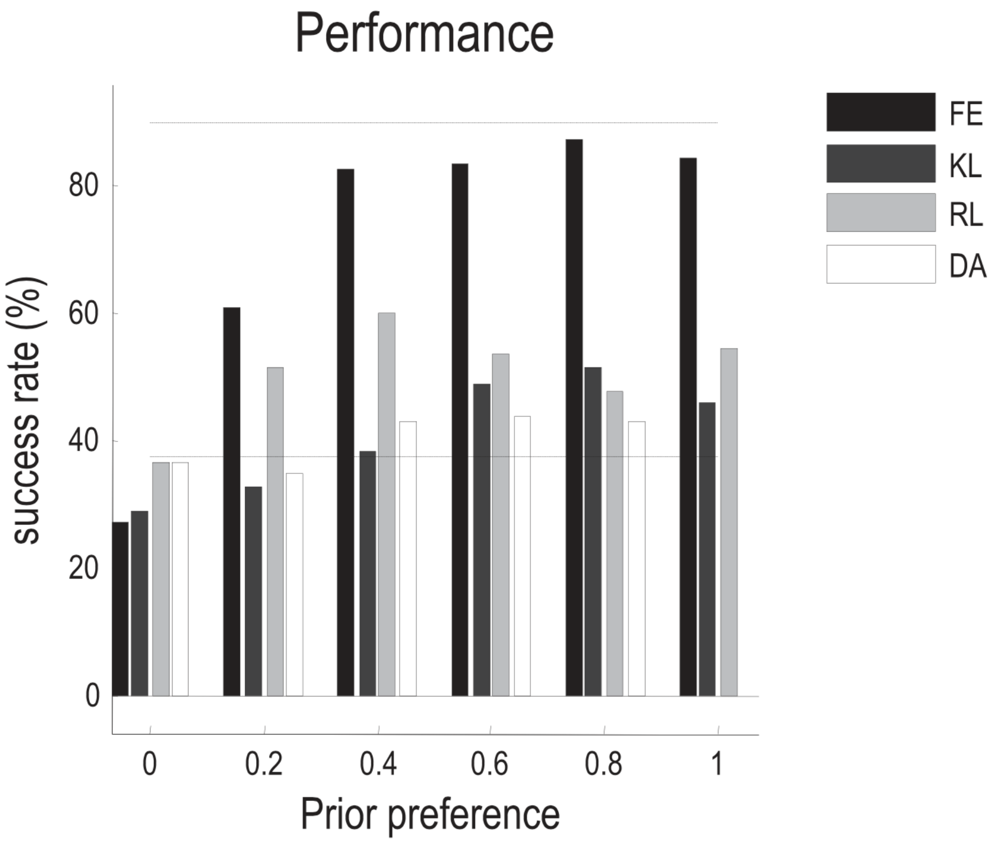


- Outcomes



$$\ln P(o_t) = U = [0, c, -c, c, -c, 0, 0]^T$$

Simulation



$$\mathbf{Q}_\tau(\pi) = E_{Q(o_\tau, s_\tau|\pi)}[\underbrace{\ln P(o_\tau|s_\tau) - \underbrace{\ln Q(o_\tau|\tilde{u}) + \underbrace{\ln P(o_\tau|m)}_{\text{Expected utility}}}_{\text{KL control}}}_{\text{Expected Free energy}}]$$

Another usage

- A simpler usage of active inference as a belief update mechanism
- Belief update by free energy minimization
- Case of a bounded rational agent

- Free energy functional with a rationality index

Complexity — Accuracy

$$\Delta F[q] = \frac{1}{\alpha}\sum_h q(h)\log\frac{q(h)}{p_0(h)} - \sum_h q(h)\log p(y \mid h)$$

- α: **Rationality Index (Bound)**
- $q(h)$: Latent variable
- $p_0(h)$: Prior
- $p(y \mid h)$: Likelihood model, Log likelihood

* Ortega et al. "Thermodynamics as a theory of decision-making with information-processing costs" (2013)
* Friston et al. "The anatomy of choice: active inference and agency" (2013)

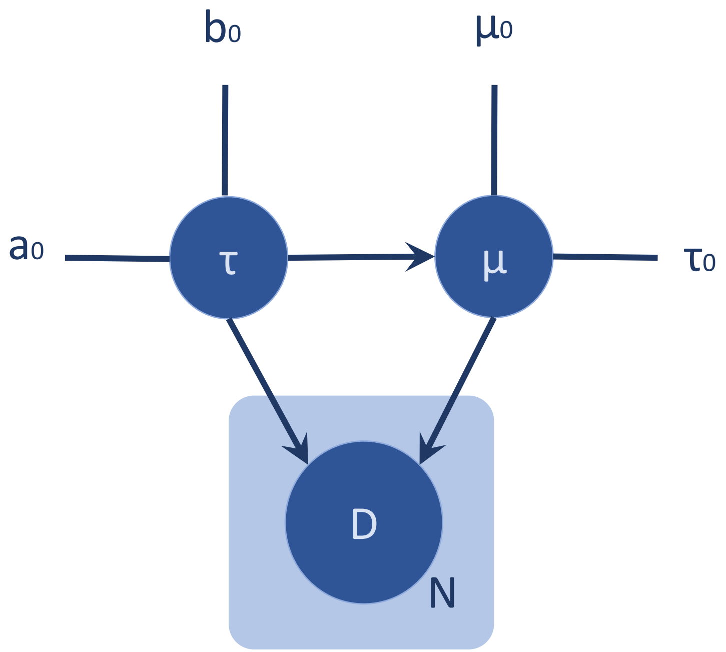
- Testing with 1-D Gaussian model
  - Observed data

$$p(D \mid \mu,\tau) = \frac{\tau^{\frac{N}{2}}}{2\pi} e^{-\frac{\tau}{2}\sum_{i=1}^{N}(x_i-\mu)^2} \sim N(\mu,\tau)$$

  - Priors

$$p(\mu \mid \tau) = \frac{\lambda_0 \tau}{2\pi} e^{-\frac{\lambda_0 \tau}{2}(\mu-\mu_0)^2} \sim N(\mu_0,\lambda_0\tau)$$

$$p(\tau) = \Gamma(a_0,b_0) = \frac{b_0 a_0}{\Gamma(a_0)} x^{a_0-1} e^{-b_0\tau} \sim \Gamma(a_0,b_0)$$

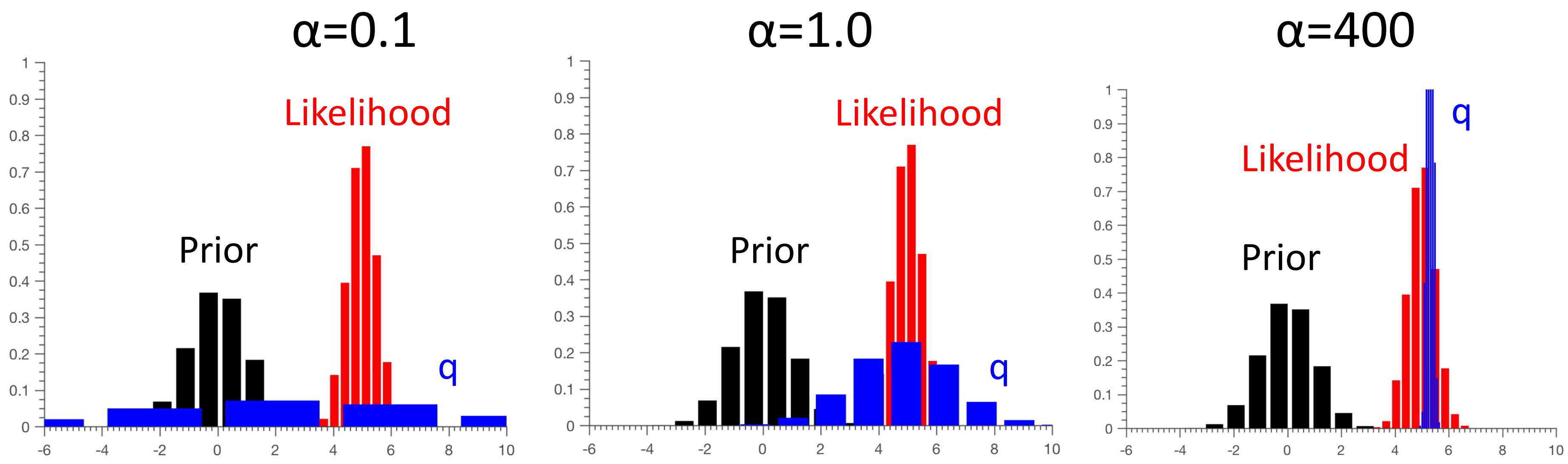


- Testing with 1-D Gaussian model
  - Free energy functional

Rationality Index:
Low α: large constraint
High α: low constraint

$$F[q(\mu,\tau)] = -\int q(\mu,\tau)\ln p(D \mid \mu,\tau) + \frac{1}{\alpha}\int q(\mu,\tau)\ln\frac{q(\mu,\tau)}{p(\mu \mid \tau)p(\tau)}$$

  - Example



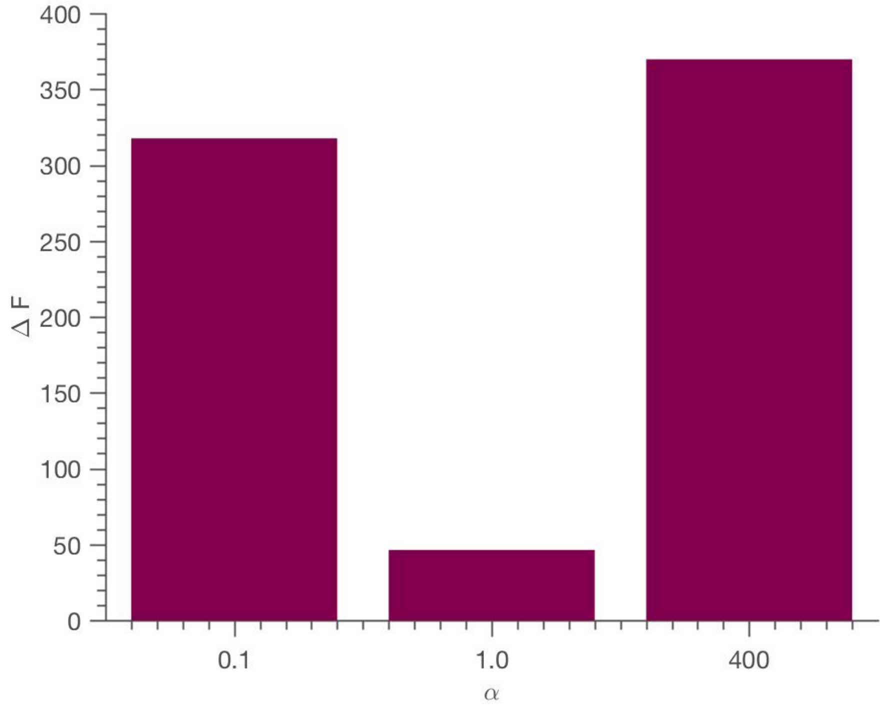
- Testing with 1-D G
  - Free energy fu

$$F[q(\mu,\tau)]$$

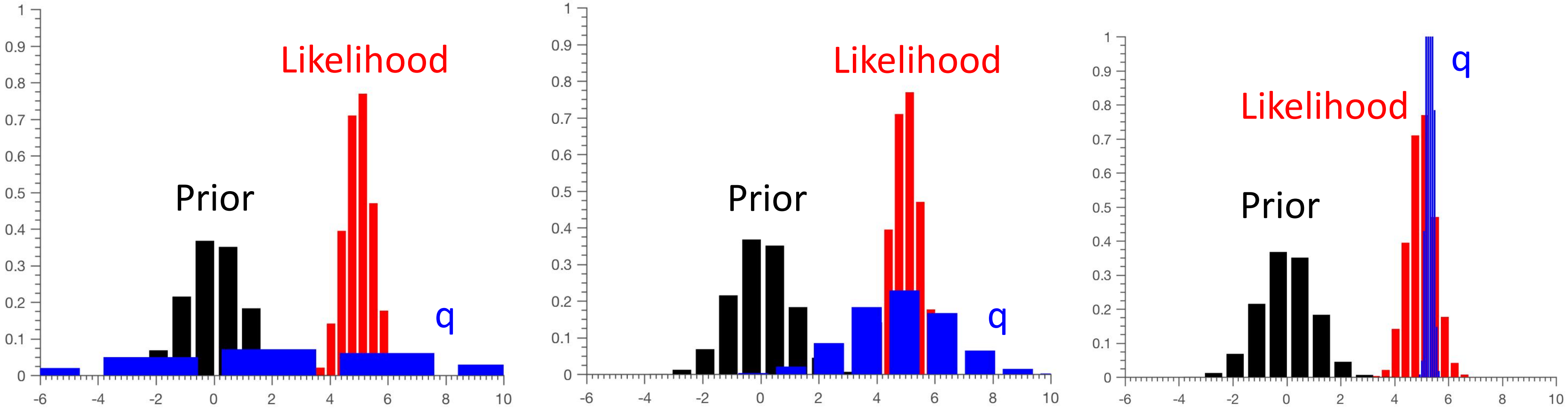
  - Example

ty Index:
rge constraint
ow constraint

$$\frac{q(\mu,\tau)}{\mu\,|\,\tau)p(\tau)}$$



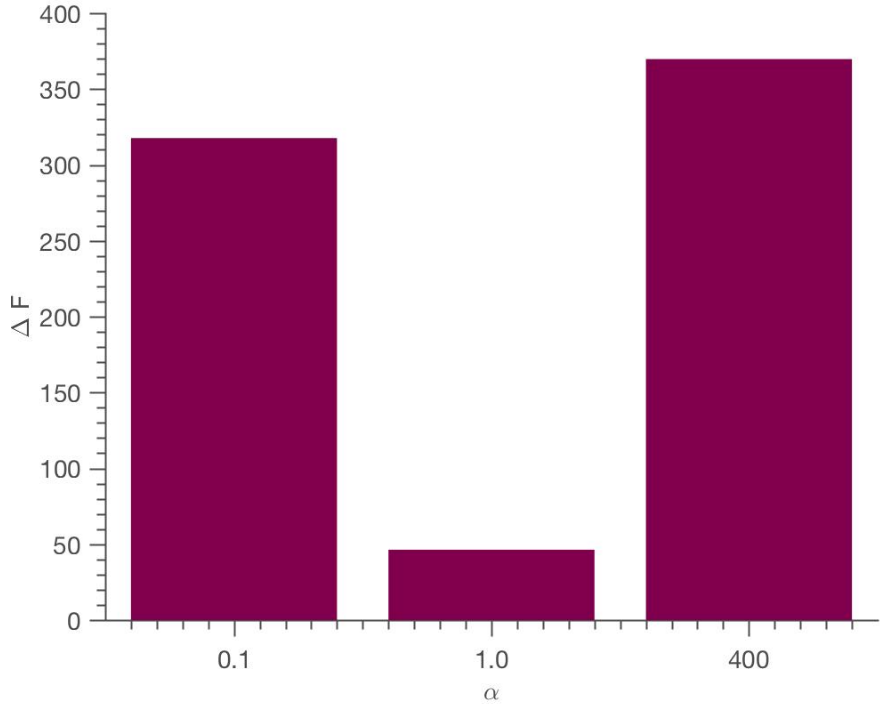
α=0.1 α=1.0 α=400



- Testing with 1-D G
  - Free energy fu

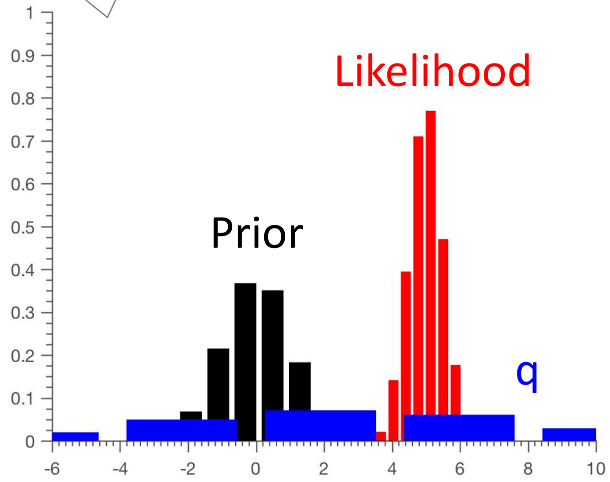
ty Index:
rge constraint
ow constraint

- Random action
- Helplessness
- Anhedonia

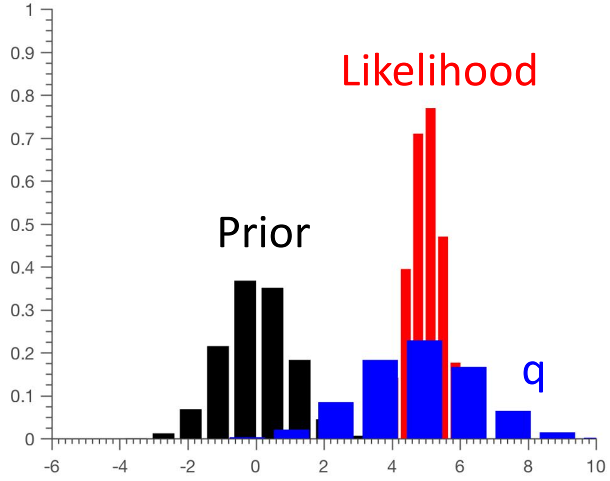


- Risk seeking
- Impulsivity
- Illusory pattern perception

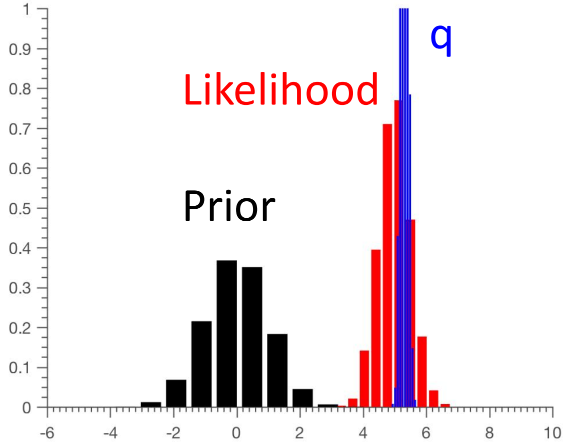
α=0.1



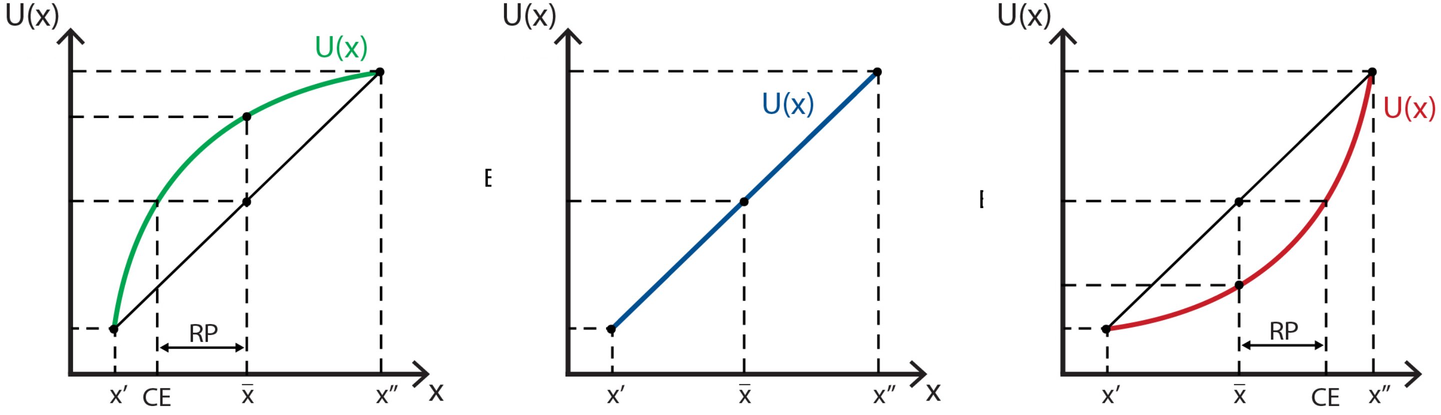
α=1.0



α=400



Agent with risk attitudes

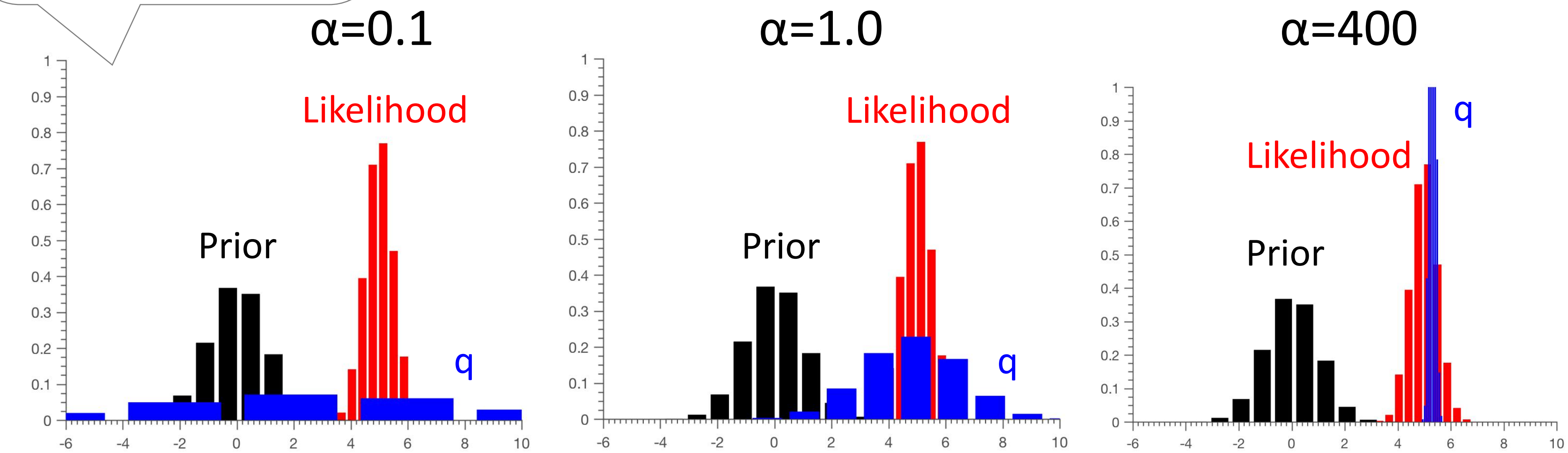


- Testing wit
  - Free en

raint
aint

- Rando
- Helple
- Anhedonia

tern perception



Calculation of the FE functional

$$F[q(\mu,\tau)] = -\int q(\mu,\tau)\ln(D|\mu,\tau) + \frac{1}{\alpha}\int q(\mu,\tau)\ln\left\{\frac{q(\mu,\tau)}{p(\mu|\tau)p(\tau)}\right\}$$

$$= -\int q(\mu,\tau)\ln(D|\mu,\tau) + \frac{1}{\alpha}\int q(\mu,\tau)\ln q(\mu) + \frac{1}{\alpha}\int q(\mu,\tau)\ln q(\tau) - \frac{1}{\alpha}\int q(\mu,\tau)\ln p(\mu|\tau) - \frac{1}{\alpha}\int q(\mu,\tau)\ln p(\tau)$$

Term 1 **Term 2** **Term 3** **Term 4** **Term 5**

Term 1 $= \int q(\mu,\tau)\ln(D|\mu,\tau) = <\ln(D|\mu,\tau)>_q$

$$= \frac{N}{2}\left[(\psi(a_N) - \ln b_N) - 2\pi\right] - \frac{\frac{a_N}{b_N}}{2}\left(\sum_1^N (x_n^2 - 2\mu_N x_n + \mu_N^2 + \frac{1}{\lambda_N}\right)$$

Expectation of log likelihood under approximate posterior

Term 2 $= \frac{1}{\alpha}\int q(\mu,\tau)\ln q(\mu) = <\ln q(\mu)>_q$

$$= \frac{1}{\alpha}\left(\frac{1}{2}\ln\frac{2\pi}{\lambda_N} + \frac{1}{2}\right)$$

Entropy of approximate posterior over the mean.

Calculation of the FE functional

Term 3

$$= \frac{1}{\alpha}\int q(\mu,\tau)\ln q(\tau) \;=< \ln q(\tau) >_q$$

$$= \frac{1}{\alpha}\left(a_N - \ln b_N + \ln \Gamma(a_N) + (1-a_N)\psi(a_N)\right)$$

Entropy of approximate posterior over the precision.

Term 4

$$= \frac{1}{\alpha}\int q(\mu,\tau)\ln p(\mu|\tau) \;=< \ln p(\mu|\tau) >_q$$

Expectation of prior on the mean over approximate posterior.

$$= \frac{1}{\alpha}\left(\frac{N}{2}\left[\ln \lambda_0 + \left(\psi(a_N) - \ln b_N\right) - 2\pi\right] - \frac{\lambda_0 \tau}{2}\left(\mu_n^2 - 2\mu_N\mu + \mu_N^2 + \frac{1}{\lambda_N}\right)\right)$$

Term 5

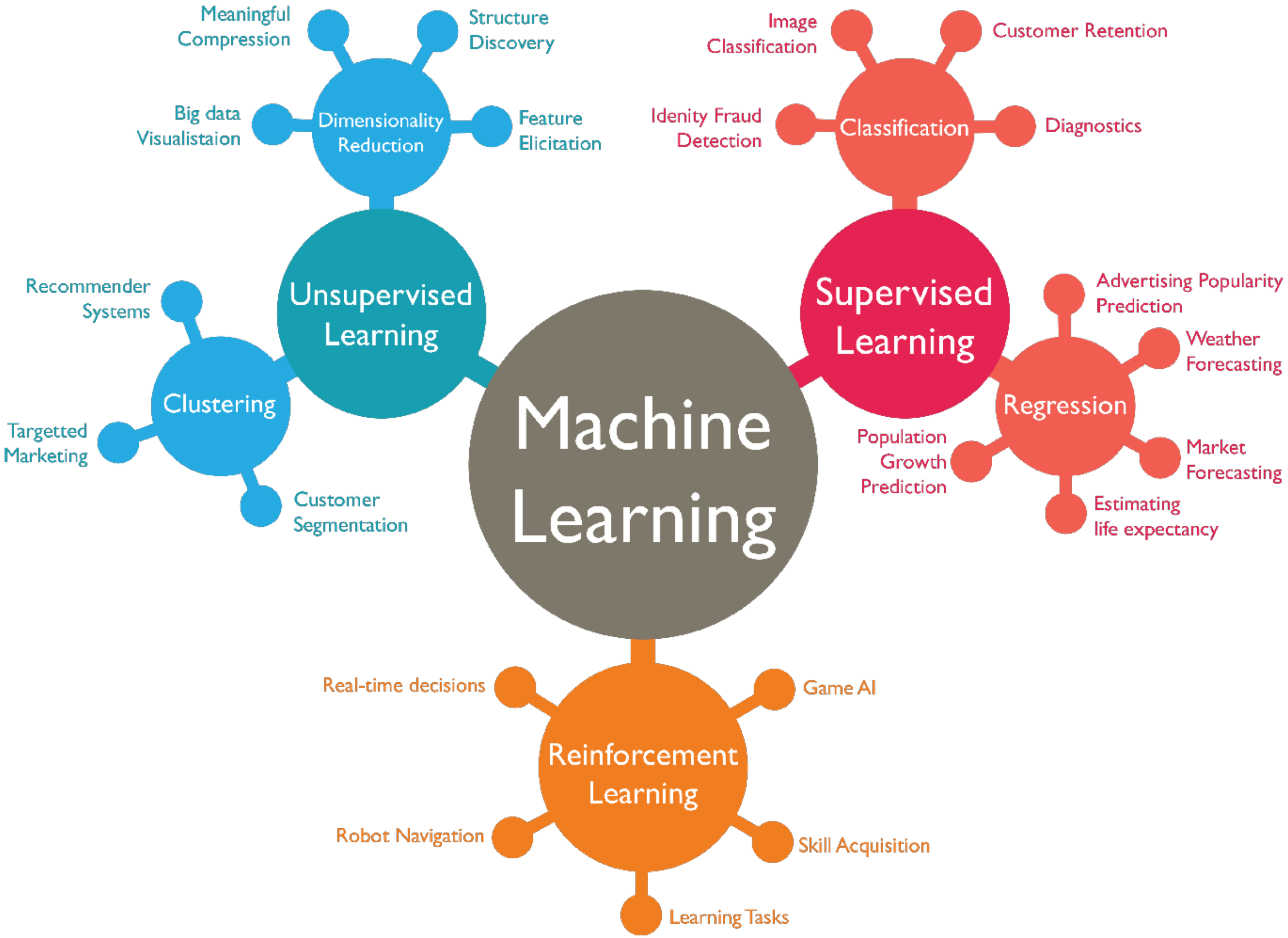
$$= \frac{1}{\alpha}\int q(\mu,\tau)\ln p(\tau) \;=< ln(p(\tau) >_q$$

Expectation of prior on precision over approximate posterior.

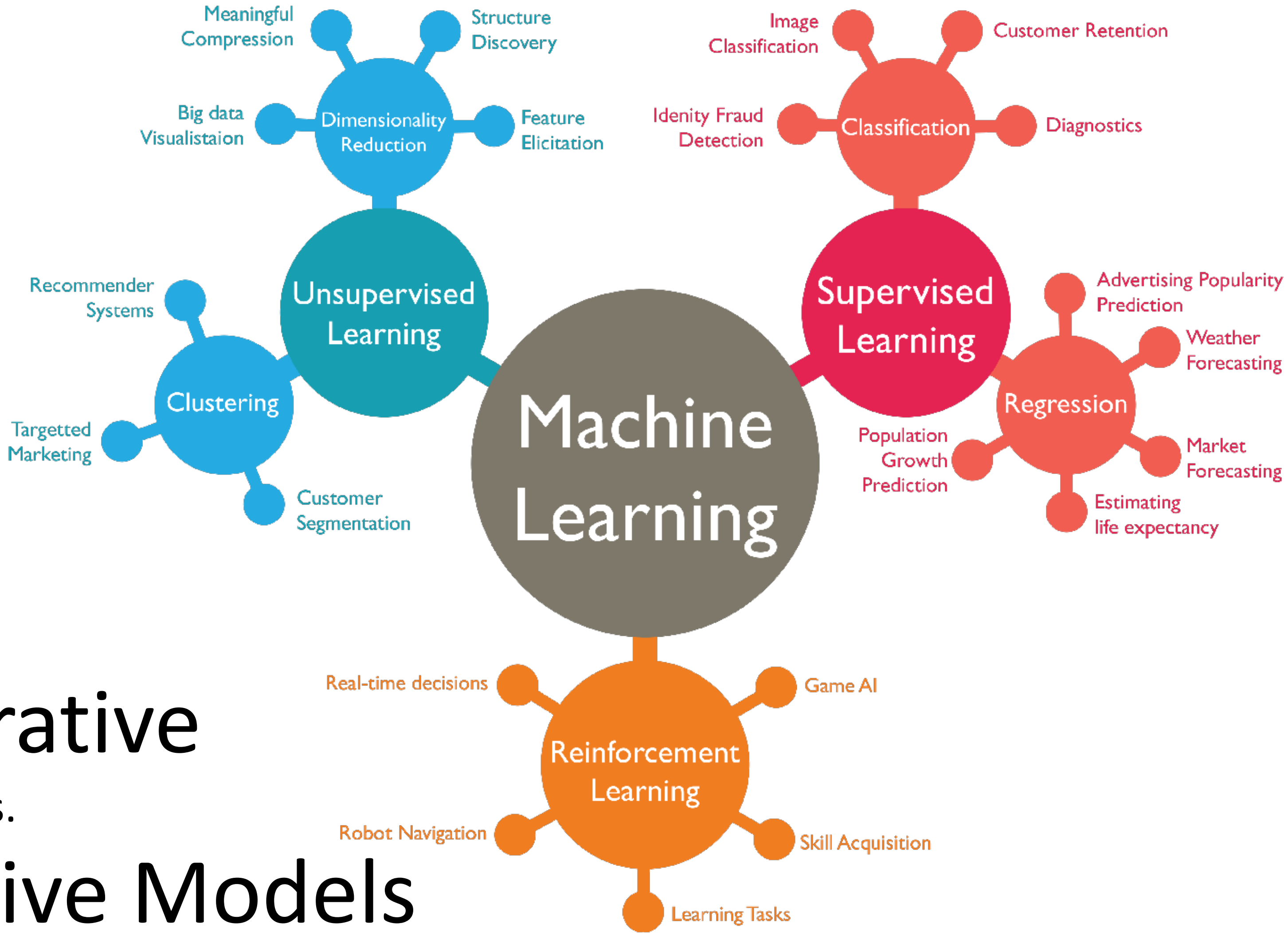
$$= \frac{1}{\alpha}\left(a_0 \ln b_0 - \ln \Gamma(a_0) + (a_0 - 1)\left(\psi(a_N) - \ln b_N - b_0\left(\frac{a_N}{b_N}\right)\right)\right)$$

Questions & Comments?

Generalities



Generalities



Generative

vs.

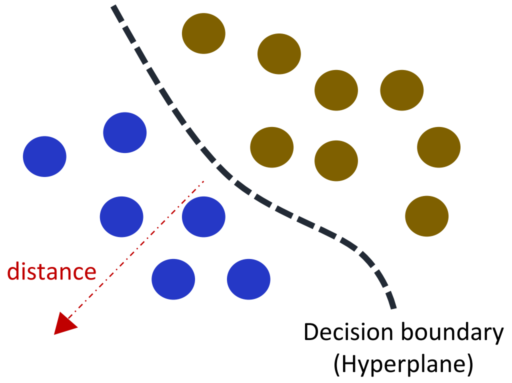
Discriminative Models

Generative vs. discriminative models

Data X
Label Y

Discriminative models learn the (hard or soft) boundary between classes

Learn $P(y|x)$ directly

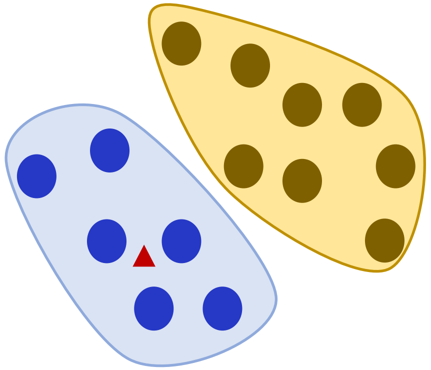


Discriminative

- Logistic regression
- SVM
- NN

Generative models model the distribution of individual classes

Model $P(x|y)$ and $P(y)$, and learn $P(y|x)$ indirectly: $P(y|x) \propto P(x|y)P(y)$



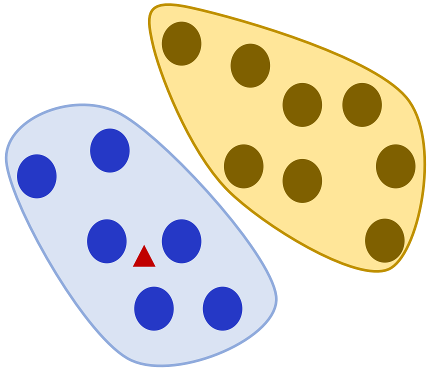
Generative

- Naïve Bayes
- Gaussian Discriminant Analysis

Generative vs. discriminative models

Data X
Label Y

→ Model $P(x|y)$ and $P(y)$, and learn $P(y|x)$ indirectly: $P(y|x) \propto P(x|y)P(y)$



Generative

- Naïve Bayes
- Gaussian Discriminant Analysis

Provides **a probability distribution for each class** in the classification problem. This give us an idea of how the data is generated. It relies heavily on Bayes rule to define, update the **prior** and derive the **posterior**.

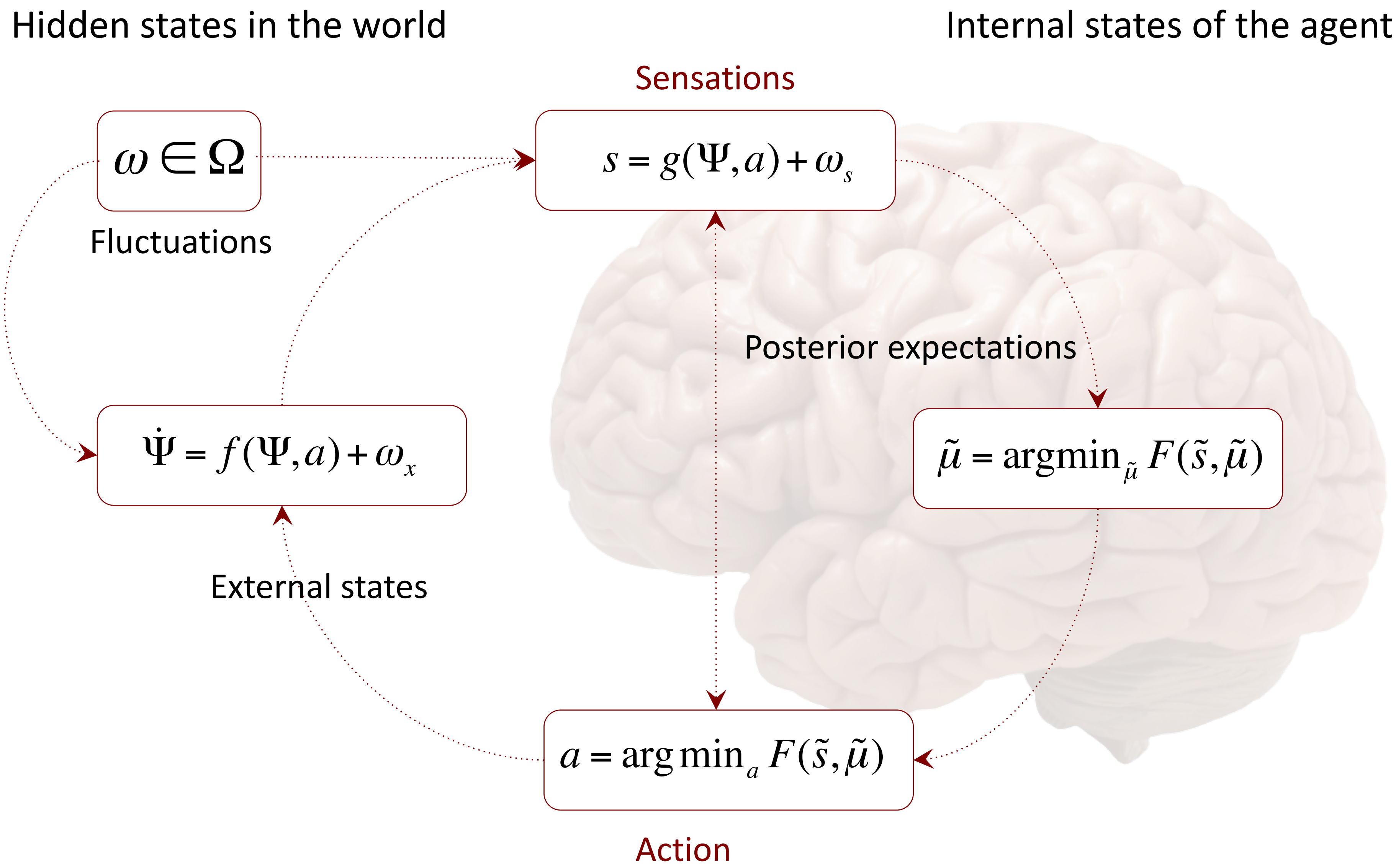
Posterior — likelihood — Prior

$$P(\theta|y) = \frac{P(y|\theta)P(\theta)}{P(y)}$$

Evidence

1. Formulation of a generative model
   - Likelihood $P(y|\theta)$
   - Prior distribution $P(\theta)$
2. Observation of data: y
3. Update of beliefs upon observations given a prior state of knowledge:

$$P(\theta|y) \propto P(y|\theta)P(\theta)$$


Hidden states in the world
Internal states of the agent
Sensations
$\omega \in \Omega$
$s = g(\Psi, a) + \omega_s$
Fluctuations
Posterior expectations
$\dot{\Psi} = f(\Psi, a) + \omega_x$
$\tilde{\mu} = \arg\min_{\tilde{\mu}} F(\tilde{s}, \tilde{\mu})$
External states
$a = \arg\min_a F(\tilde{s}, \tilde{\mu})$
Action