The physics of sentience

Karl Friston

Abstract: how can we understand ourselves as sentient creatures? And what are the principles that underwrite sentient behavior? This presentation uses the free energy principle to furnish an account in terms of active inference. First, we will try to understand sentience from the point of view of physics; in particular, the properties that self-organizing systems—that distinguish themselves from their lived world—must possess. This formulation is based on the following arguments: if a system can be differentiated from its external milieu, then its internal and external states must be conditionally independent. Crucially, this independence equips internal states with an information geometry, pertaining to probabilistic beliefs about something; namely external states. In short, internal states will appear to infer—and act on—their world to preserve their integrity. This leads to a Bayesian mechanics, which can be neatly summarized as self-evidencing. In the second half of the talk, we will unpack these ideas using constructs from neurobiology — and simulations of Bayesian belief updating in the brain.

Key words: *active inference · autopoiesis · cognitive · dynamics · free energy · epistemic value · self-organization.*

.

I'd like to meet Helmholtz. He believed in unconscious perceptual inference. He thought that vision involves many inferences: one infers, from the proximal stimulus, the state of the world that generated it.

Another aspect of Helmholtz's work concerned free energy. What he didn't know was that to infer a complex model, a statistician would advise you to do the inference by determining the best way for your model to explain the data. A model can have multiple ways of explaining the data. You need to determine the probability of each of these ways. Variational inference approximates the probabilities while ensuring model improvement.

Thus, the two completely different aspects of Helmholtz's work—free energy and perceptual inference—are closely related. Free energy is the key to perceptual inference. And I'd like to tell Helmholtz this.

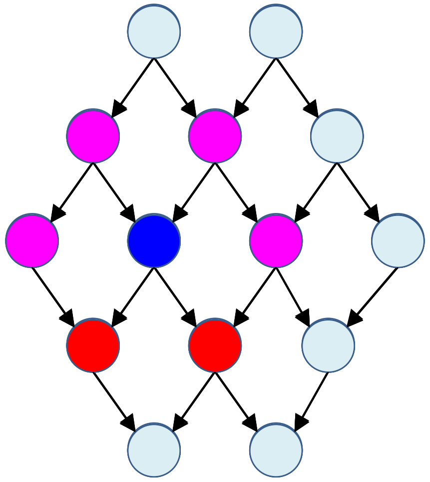
The statistics of life	Markov blankets and Bayesian mechanics
The anatomy of inference	predictive coding and neuronal networks
Action and perception	active inference and agency

"How can the events in space and time which take place within the spatial boundary of a living organism be accounted for by physics and chemistry?"

(Erwin Schrödinger 1943)

The Markov blanket as a statistical boundary

(parents, children and parents of children)



η External states

μ Internal states

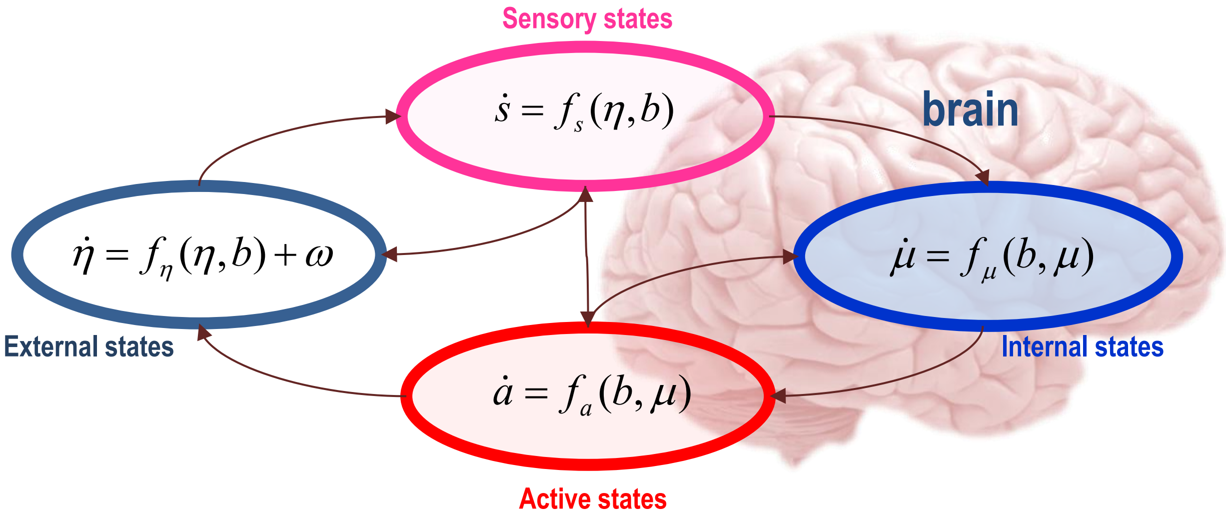
s Sensory states

a Active states

} b Blanket states (sensory and active states)

} π Particular states (internal and blanket states)

Markov blankets



lemma: *any random dynamical system that possesses a Markov blanket (m) will appear to self-evidence*

$$\dot{x} = f(x) + \omega$$

$$p(x \mid m)$$

Density dynamics $\dot{p}(x \mid m) = \nabla \cdot (\Gamma \nabla - f)p$

And its solution in terms of a Helmholtz decomposition

$$\dot{p}(x \mid m) = 0 \Leftrightarrow f(x) = (\Gamma - Q)\nabla \ln p(x \mid m)$$

The dynamics of self-organisation (to nonequilibrium steady-state)

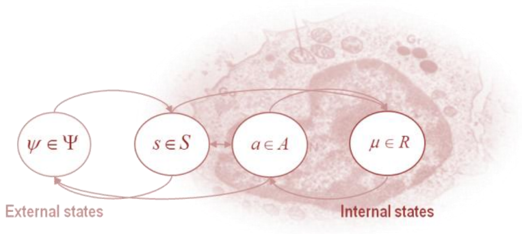
$$f_Q = -Q \cdot \nabla \ln p(x \mid m)$$

$$f_\Gamma = \Gamma \cdot \nabla \ln p(x \mid m)$$

$$f = f_\Gamma + f_Q = (\Gamma - Q)\nabla \ln p(x \mid m)$$

The dynamics (i.e., flow) at NESS

But what about the Markov blanket?

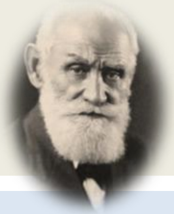


$$\dot{\boldsymbol{\mu}} = (\Gamma - Q)\nabla_{\mu} \ln p(s \mid m)$$ **Perception**

$$\dot{\mathbf{a}} = (\Gamma - Q)\nabla_{a} \ln p(s \mid m)$$ **Action**

$-F(s, \boldsymbol{\mu}) = \ln p(s \mid m) =$ **Value**

Reinforcement learning
Optimal control theory
Expected utility theory



Pavlov

$F(s, \boldsymbol{\mu}) = -\ln p(s \mid m) =$ **Surprise**

Infomax principle
Minimum redundancy
Free-energy principle

Barlow

$\mathrm{E}\left[F(s, \boldsymbol{\mu})\right] = \mathrm{H}[p(s \mid m)] =$ **Entropy**

Self-organization
Synergetics
Homoeostasis

Haken

$p(s \mid m) =$ **Evidence**

Bayesian brain
Evidence accumulation
Predictive coding

Helmholtz

The statistics of life — Markov blankets and Bayesian mechanics

The anatomy of inference — **predictive coding and neuronal networks**

Action and perception — active inference and epistemic affordance

Giuseppe Arcimboldo, **The Vegetable Gardener** (c.1590). Oil on panel. Our percepts are constrained by what we expect to see and the hypotheses that can be called upon to explain our sensory input. Arcimboldo, "a 16th century Milanese artist who was a favorite of the Viennese, illustrates this dramatically by using fruits and vegetables to create faces in his paintings.

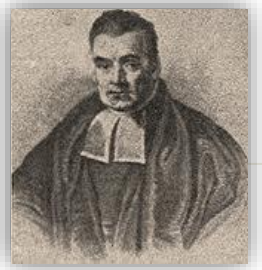
Hermann von Helmholtz

“Objects are always imagined as being present in the field of vision as would have to be there in order to produce the same impression on the nervous mechanism” - *von Helmholtz*

Richard Gregory

Geoffrey Hinton

Peter Dayan



Thomas Bayes

The Helmholtz machine and the Bayesian brain

Richard Feynman

Hermann von Helmholtz

“Objects are always imagined as being present in the field of vision as would have to be there in order to produce the same impression on the nervous mechanism” - *von Helmholtz*

Richard Gregory

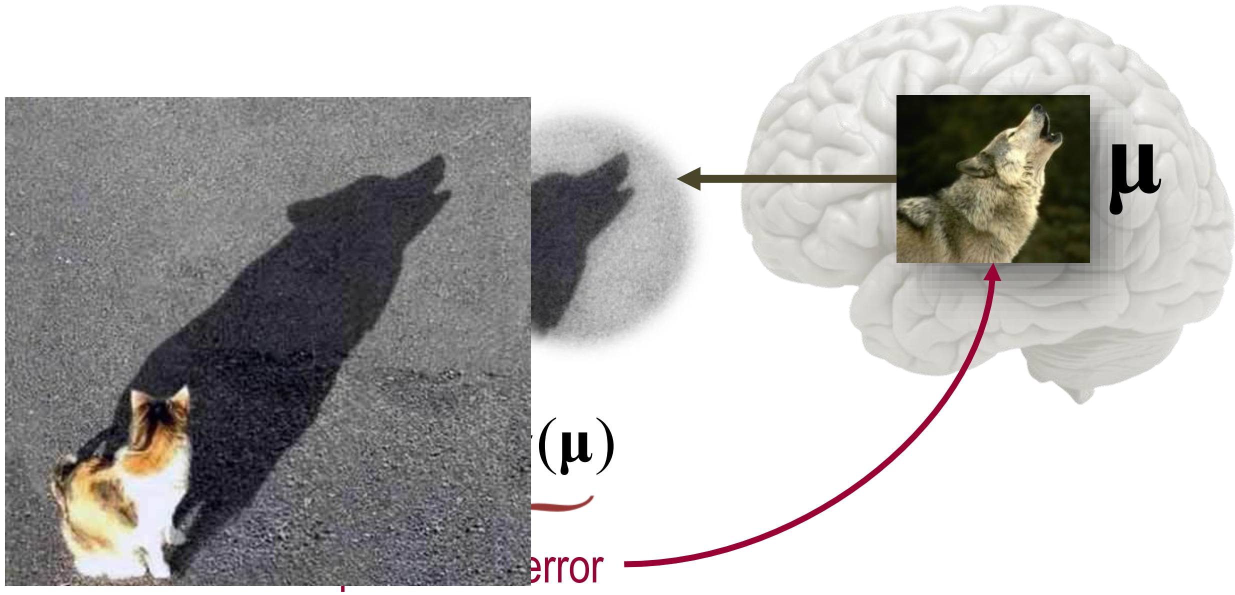
Impressions on the Markov blanket…

The Helmholtz decomposition, Bayesian filtering and predictive coding

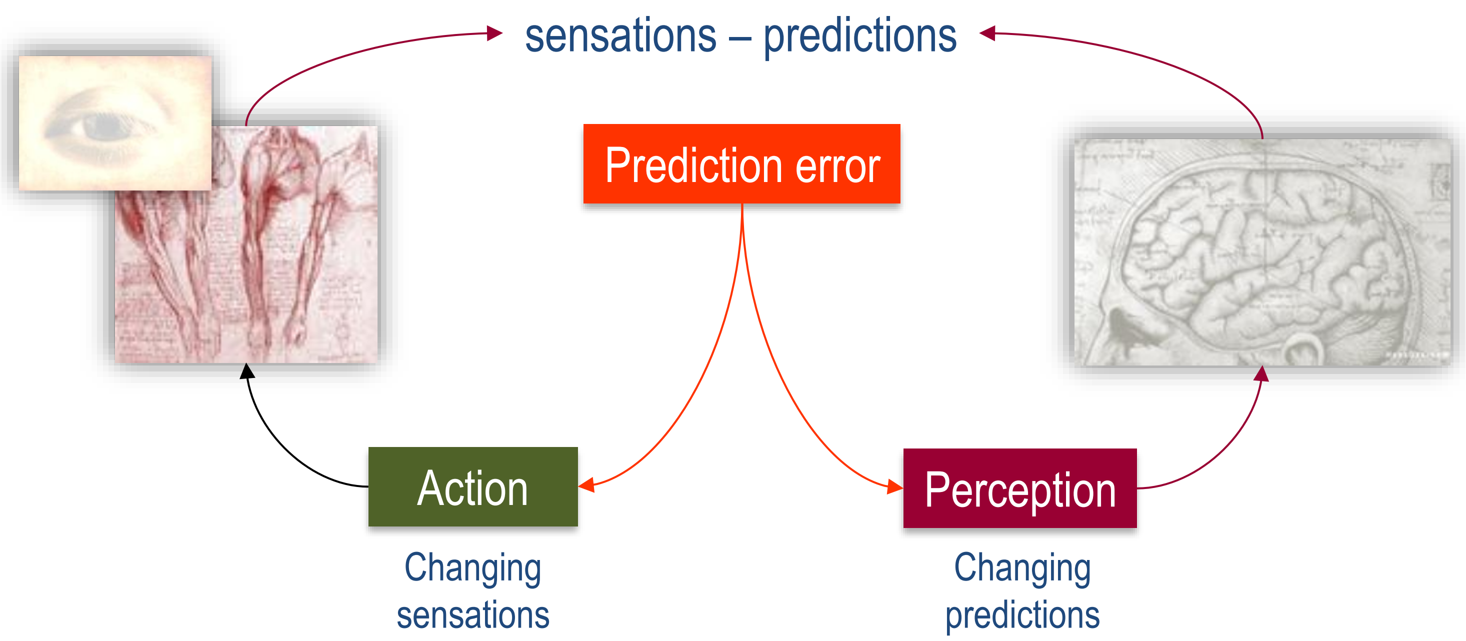
$$q_{\boldsymbol{\mu}}(\eta) = \square\,(\boldsymbol{\mu}, \Gamma(\boldsymbol{\mu}))$$

$$\begin{aligned}\dot{\boldsymbol{\mu}} &= (Q - \Gamma) \cdot \nabla F(s, \boldsymbol{\mu}) \\ &= D\boldsymbol{\mu} - \nabla_{\boldsymbol{\mu}} g \cdot \varepsilon\end{aligned}$$

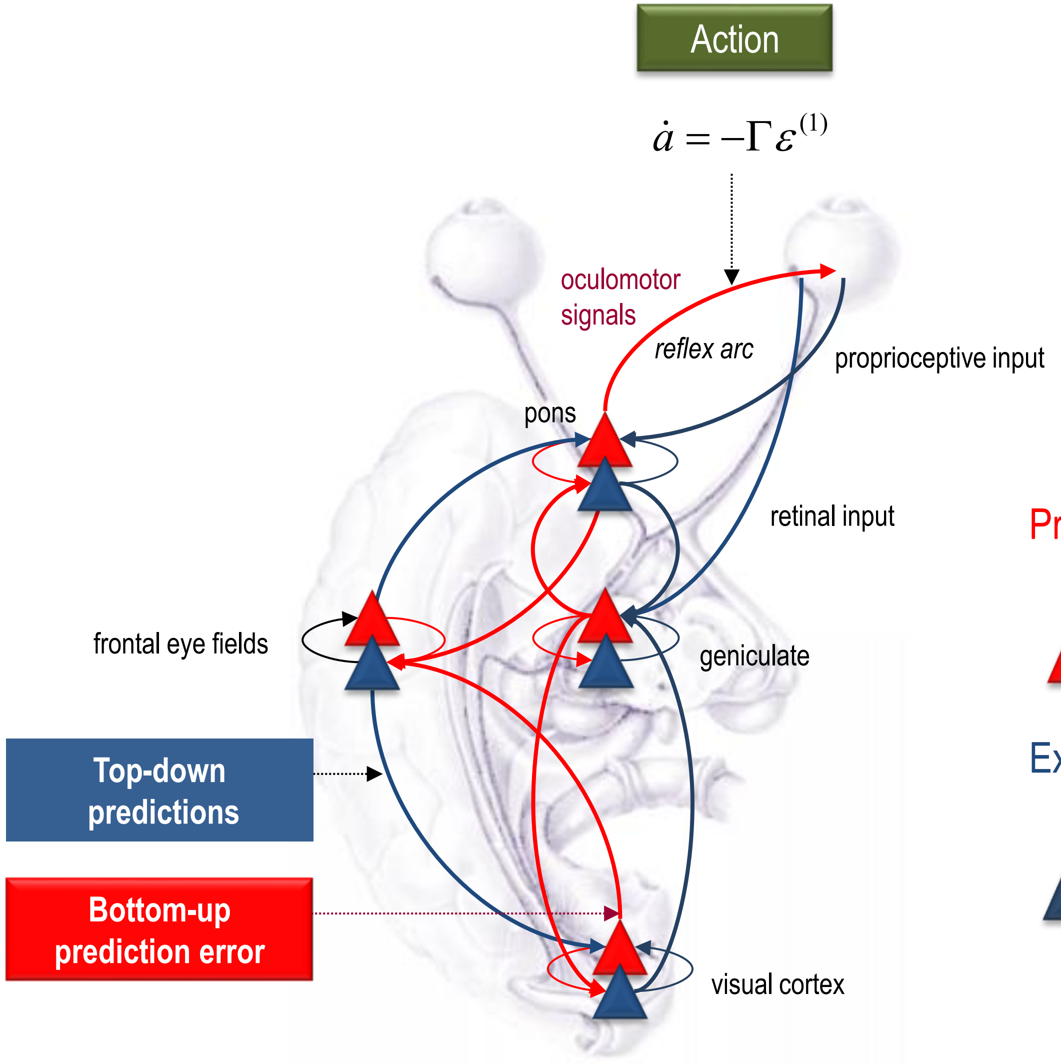
prediction update



Making our own sensations



Predictive coding with reflexes




David Mumford

Perception

Prediction error (superficial pyramidal cells)

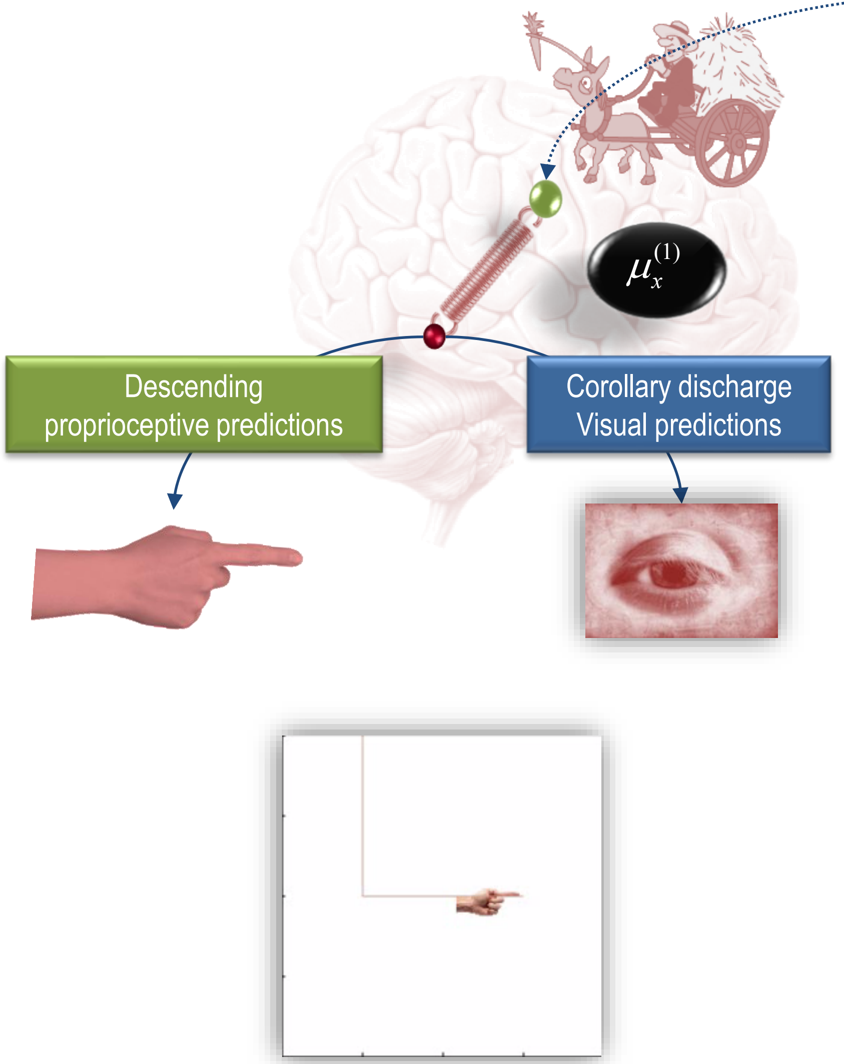
$$\varepsilon^{(i)} = \mu^{(i-1)} - g^{(i)}(\mu^{(i)})$$

Expectations (deep pyramidal cells)

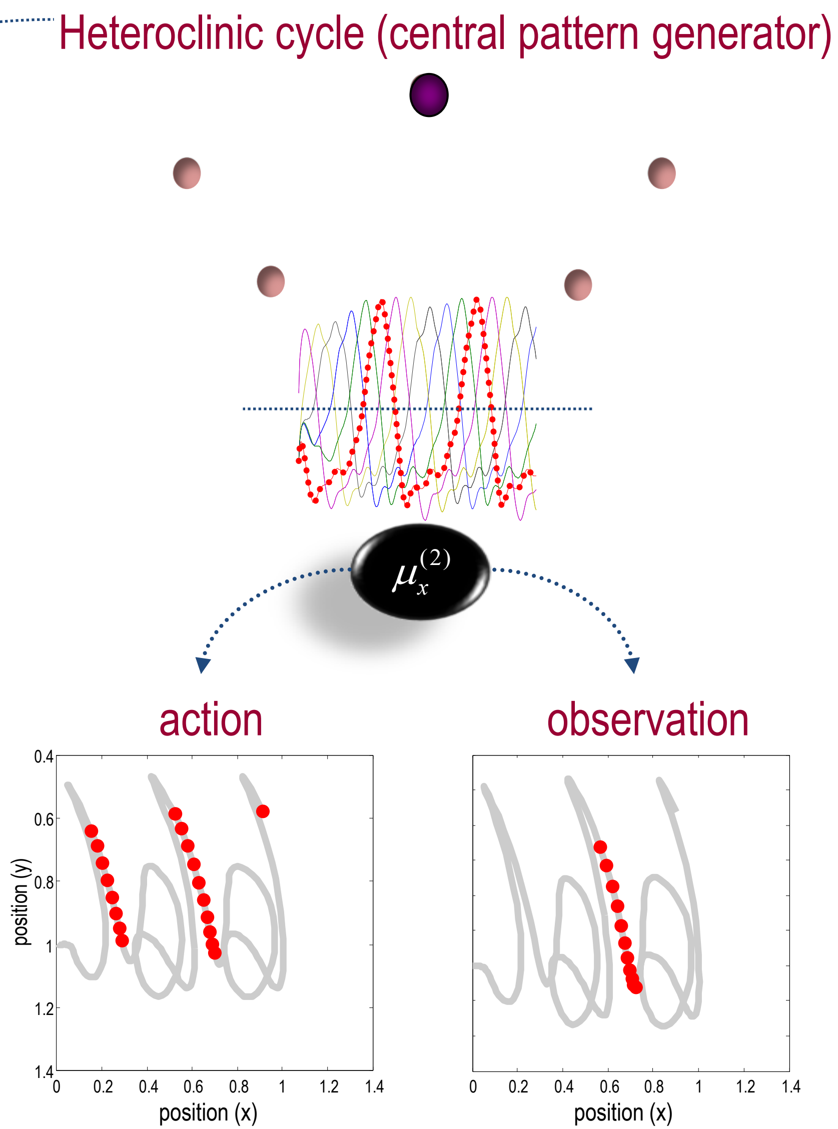
$$\dot{\mu}^{(i)} = D\mu^{(i)} - \Gamma \varepsilon^{(i)}$$

The statistics of life	Markov blankets and Bayesian mechanics
The anatomy of inference	predictive coding and neuronal networks
Action and perception	**active inference and epistemic affordance**

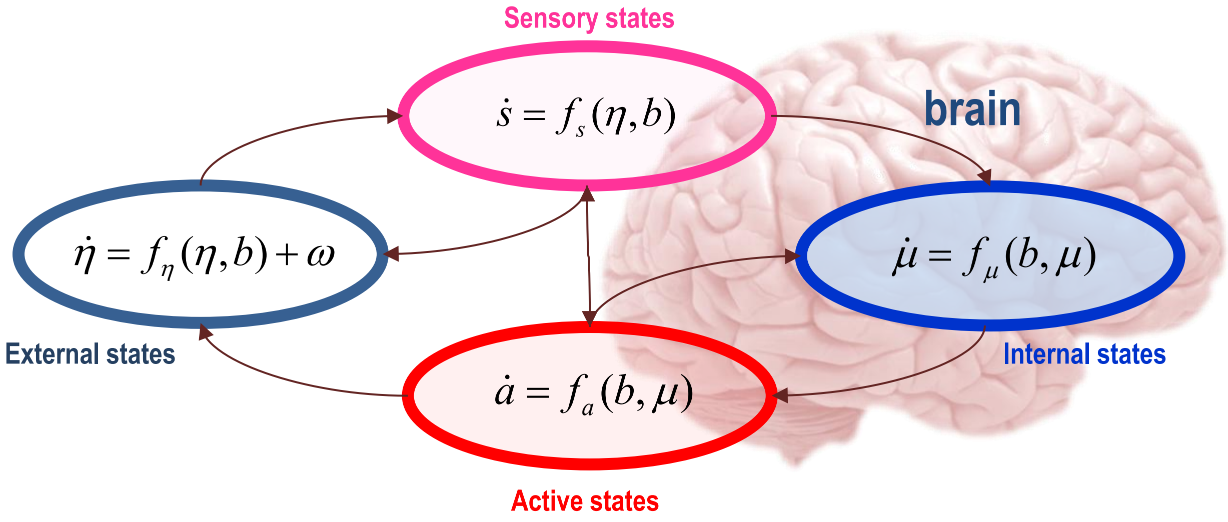
Active inference

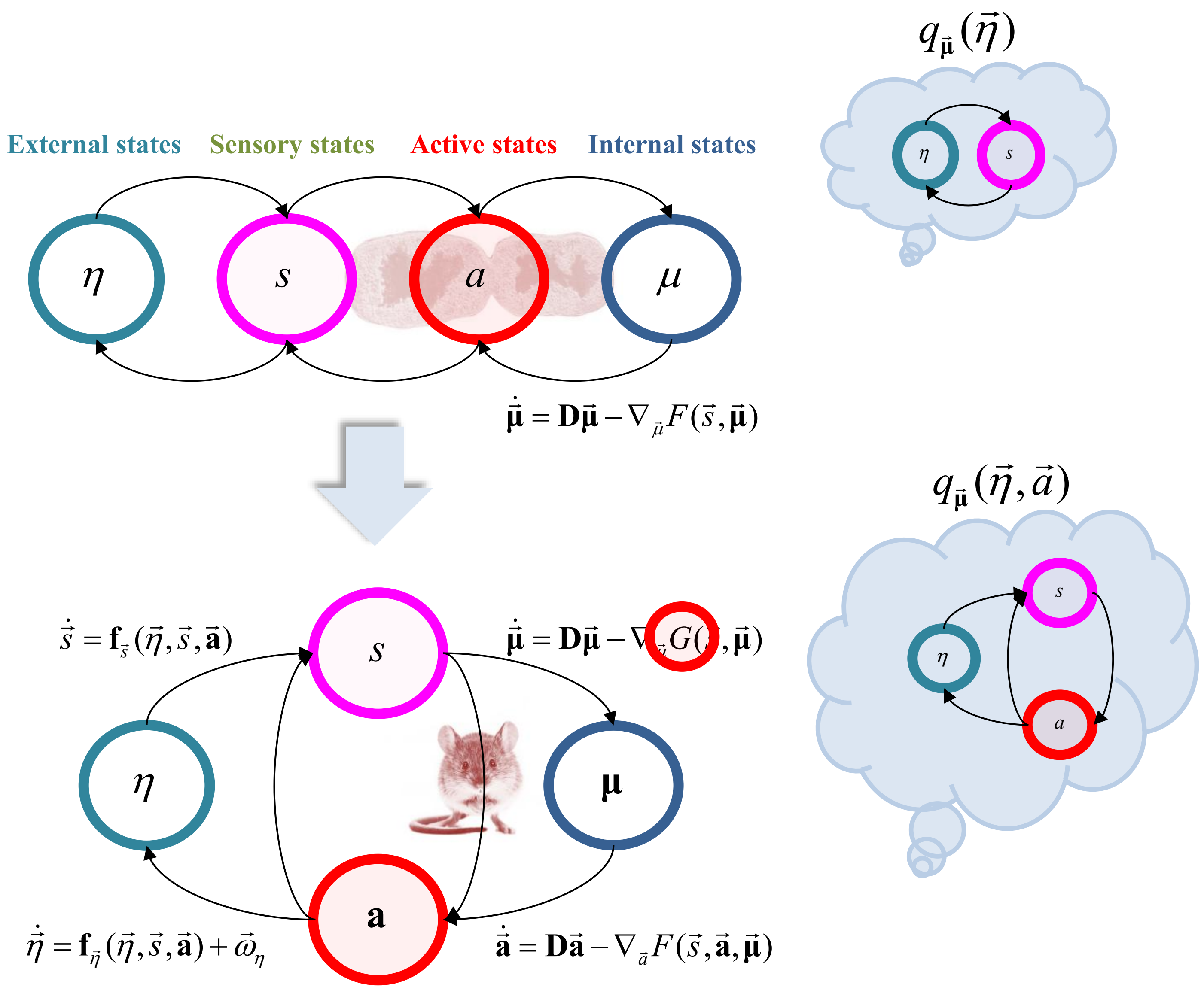


$$\dot{a} = -\nabla_a \tilde{s} \cdot \Pi \cdot \tilde{\varepsilon}^{(1)}$$



Markov blankets





External states
Sensory states
Active states
Internal states
η
s
a
μ
$\dot{\vec{\mu}} = \mathbf{D}\vec{\mu} - \nabla_{\vec{\mu}} F(\vec{s}, \vec{\mu})$
$q_{\vec{\mu}}(\vec{\eta})$
$q_{\vec{\mu}}(\vec{\eta}, \vec{a})$
$\dot{\vec{s}} = \mathbf{f}_{\vec{s}}(\vec{\eta}, \vec{s}, \vec{\mathbf{a}})$
$\dot{\vec{\mu}} = \mathbf{D}\vec{\mu} - \nabla_{\vec{\mu}} G(\vec{s}, \vec{\mu})$
$\dot{\vec{\eta}} = \mathbf{f}_{\vec{\eta}}(\vec{\eta}, \vec{s}, \vec{\mathbf{a}}) + \vec{\omega}_\eta$
$\dot{\vec{\mathbf{a}}} = \mathbf{D}\vec{\mathbf{a}} - \nabla_{\vec{a}} F(\vec{s}, \vec{\mathbf{a}}, \vec{\mu})$

Planning as inference

$$F(s,a) = \mathrm{E}_{q(\eta|a)}[\overbrace{\ln q(\eta|a) - \ln p(\eta)}^{\text{Complexity}} - \overbrace{\ln p(s|\eta)}^{\text{Accuracy}}]$$

$$= \mathrm{E}_{q(\eta|a)}[\underbrace{\ln q(\eta|a) - \ln p(\eta|s)}_{\text{Divergence}} - \underbrace{\ln p(s)}_{\text{Log-evidence}}]$$

$$G(a) = \mathrm{E}_{p(s,\eta|a)}[\overbrace{\ln p(\eta|a) - \ln p(\eta)}^{\text{Risk}} \overbrace{- \ln p(s|\eta)}^{\text{Ambiguity}}]$$

$$= \mathrm{E}_{p(s,\eta|a)}[\underbrace{\ln p(\eta|a) - \ln p(\eta|s)}_{\text{Intrinsic value}} - \underbrace{\ln p(s)}_{\text{Extrinsic value}}]$$

Bayesian surprise and Infomax

No prior beliefs or preferences:

$$\mathrm{E}_p[D_{KL}[p(\eta|s,a) \| p(\eta|a)]]$$

Bayesian surprise

$$D_{KL}[p(\eta,s|a) \| p(\eta|a)p(s|a)]$$

Mutual information

KL or risk-sensitive control

No ambiguity (i.e., known states):

$$D_{KL}[p(\eta|a) \| p(\eta)]$$

$$=$$

$$D_{KL}[p(s|a) \| p(s)]$$

Divergence from prior preferences

Expected utility theory

No uncertainty or risk:

$$\mathrm{E}_q[\ln p(\eta)]$$

$$=$$

$$\mathrm{E}_q[\ln p(s)]$$

Expected value

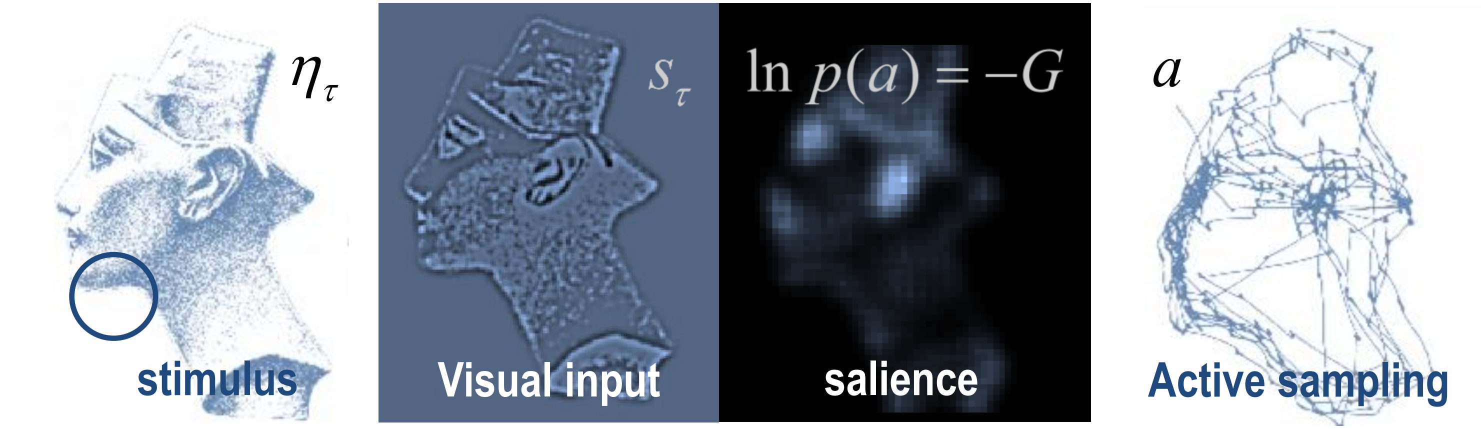
Planning as inference

$$F(s,a) = E_{q(\eta|a)}[\overbrace{\ln q(\eta|a) - \ln p(\eta)}^{\text{Complexity}} - \overbrace{\ln p(s|\eta)}^{\text{Accuracy}}]$$

$$= E_{q(\eta|a)}[\underbrace{\ln q(\eta|a) - \ln p(\eta|s)}_{\text{Divergence}} - \underbrace{\ln p(s)}_{\text{Log-evidence}}]$$

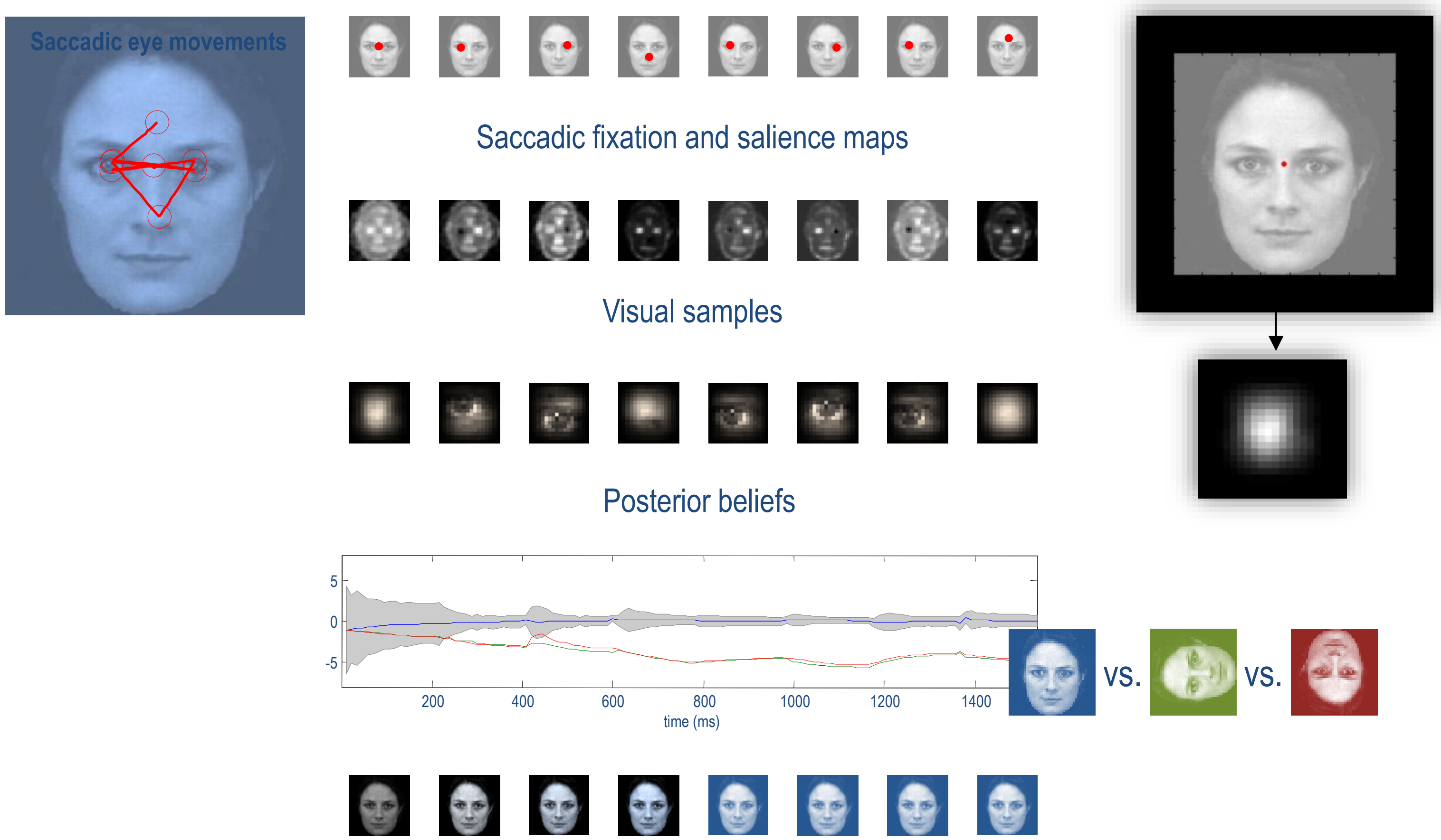
$$G(a) = E_{p(s,\eta|a)}[\overbrace{\ln p(\eta|a) - \ln p(\eta)}^{\text{Risk}} \overbrace{- \ln p(s|\eta)}^{\text{Ambiguity}}]$$

$$= E_{p(s,\eta|a)}[\underbrace{\ln p(\eta|a) - \ln p(\eta|s)}_{\text{Intrinsic value}} - \underbrace{\ln p(s)}_{\text{Extrinsic value}}]$$



Sampling the world to resolve uncertainty

Epistemic foraging and active vision



Hermann von Helmholtz

“Each movement we make by which we alter the appearance of objects should be thought of as an experiment designed to test whether we have understood correctly the invariant relations of the phenomena before us, that is, their existence in definite spatial relations.”

‘The Facts of Perception’ (1878) in The Selected Writings of Hermann von Helmholtz, Ed. R. Karl, Middletown: Wesleyan University Press, 1971 p. 384

Thank you

And thanks to collaborators:

Rick Adams
Ryszard Auksztulewicz
Andre Bastos
Sven Bestmann
Harriet Brown
Jean Daunizeau
Mark Edwards
Chris Frith
Thomas FitzGerald
Xiaosi Gu
Stefan Kiebel
James Kilner
Christoph Mathys
Jérémie Mattout
Rosalyn Moran
Dimitri Ognibene
Sasha Ondobaka
Thomas Parr
Will Penny
Giovanni Pezzulo
Lisa Quattrocki Knight
Francesco Rigoli
Klaas Stephan
Philipp Schwartenbeck

And colleagues:

Micah Allen
Felix Blankenburg
Andy Clark
Peter Dayan
Ray Dolan
Allan Hobson
Paul Fletcher
Pascal Fries
Geoffrey Hinton
James Hopkins
Jakob Hohwy
Mateus Joffily
Henry Kennedy
Simon McGregor
Read Montague
Tobias Nolte
Anil Seth
Mark Solms
Paul Verschure

And many others