MPS-UCL Symposium and Advanced Course on Computational Psychiatry and Ageing Research
Ringberg Castle, Bavaria, Germany

The Bayesian brain, free energy and psychopathology

Karl Friston
University College London

Abstract

How much about our interactions with – and experience of – our world can be deduced from basic principles? This talk reviews recent attempts to understand the self-organised behaviour of embodied agents, like ourselves, as satisfying basic imperatives for sustained exchanges with the environment. In brief, one simple driving force appears to explain many aspects of perception and action – the minimisation of surprise or prediction error. In the context of perception, this corresponds to Bayes-optimal predictive coding (that suppresses exteroceptive prediction errors) and – in the context of action – reduces to classical motor reflexes (that suppress proprioceptive prediction errors). We will look at some of the phenomena that emerge from this single principle; such as perceptual synthesis and action selection. I will focus on the key role of precision in making predictions under uncertainty. Neurobiologically, precision may be encoded by the postsynaptic gain of neuronal populations reporting prediction error and is a clear target of neuromodulatory pathologies implicated in many psychiatric disorders. I hope to illustrate this using simulations of hallucinations and failures of affordance, of the sort seen schizophrenia and Parkinson's disease.

.

.

Overview

The free-energy principle	Action and perception Generative models Predictive coding
Perception	Birdsong and attractors Deep models Simulated lesions and false inference
Action	Affordance and attractors Deep models Simulated lesions and false inference

Hermann von Helmholtz

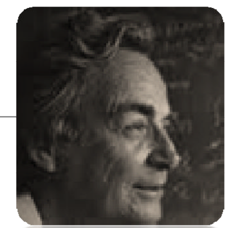
“Objects are always imagined as being present in the field of vision as would have to be there in order to produce the same impression on the nervous mechanism” - von Helmholtz

Richard Gregory

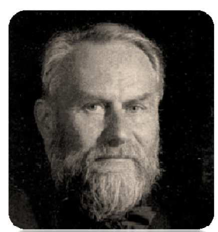
Geoffrey Hinton

Thomas Bayes

From the Helmholtz machine to the Bayesian brain and self-organization

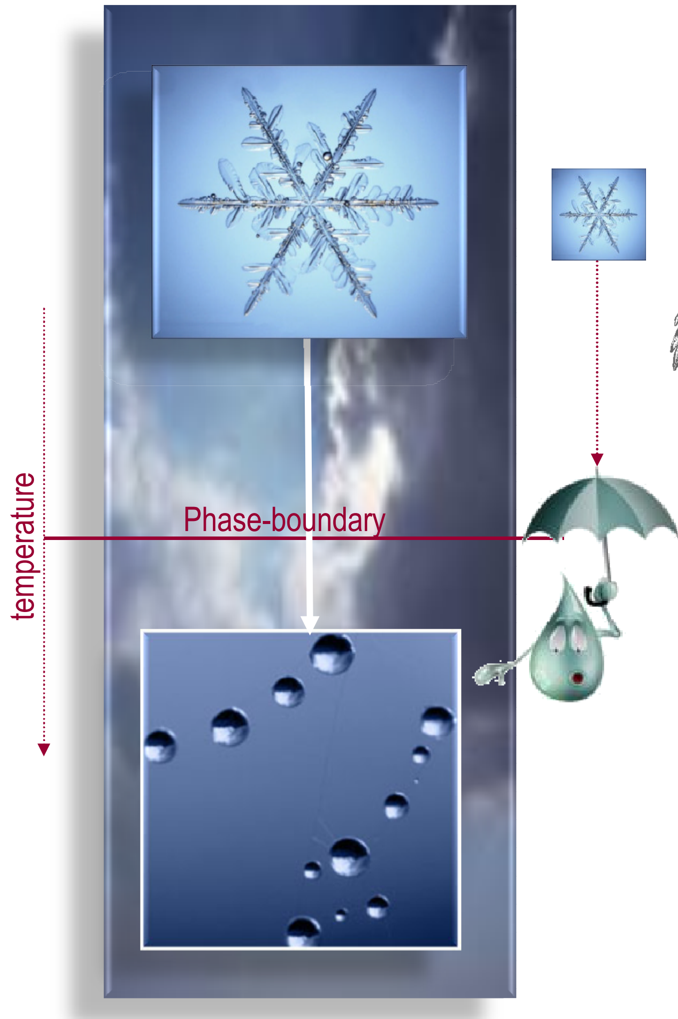


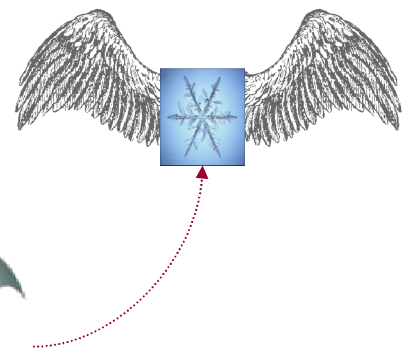
Richard Feynman



Hermann Haken

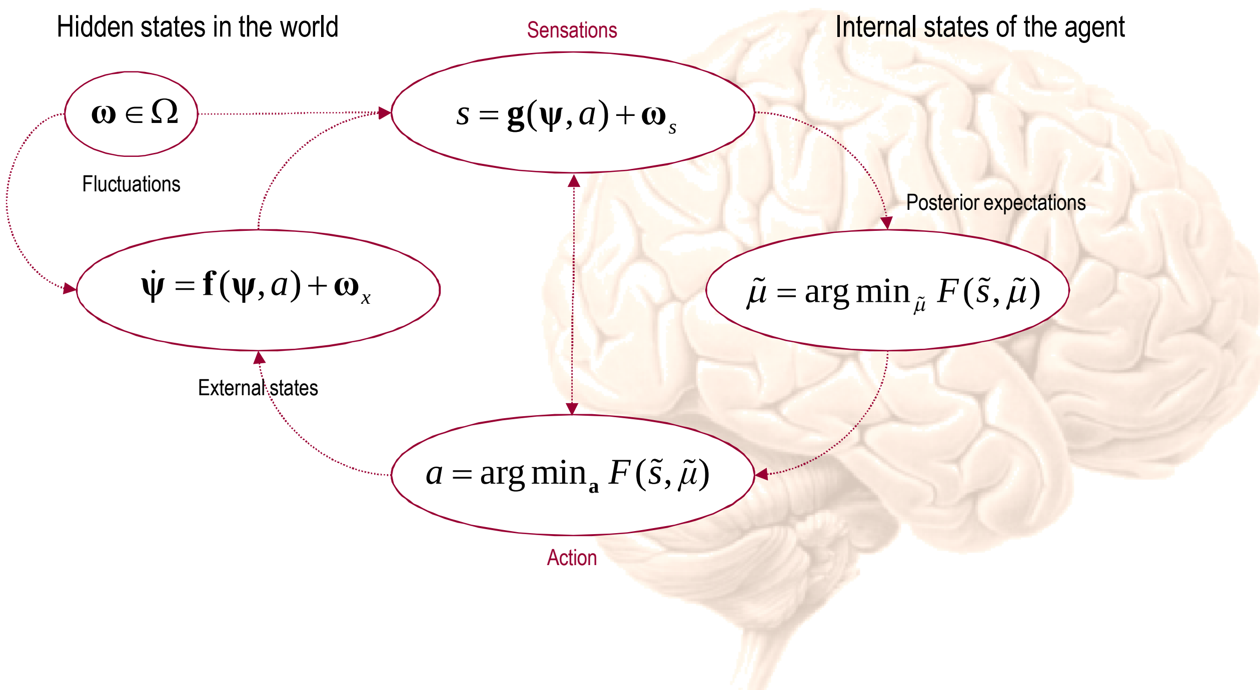
What is the difference between a snowflake and a bird?





…a bird can act (to avoid surprises)

The basic ingredients



The principle of least free energy (and minimising surprise)


$$F(\tilde{s},\mu,m) = -\ln p(\tilde{s}\,|\,m) + D_{KL}[q(\tilde{\psi}\,|\,\mu),\, p(\psi\,|\,\tilde{s})]$$

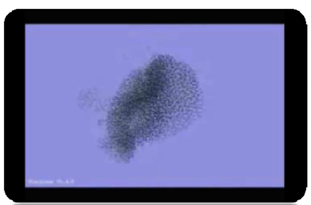
$$= E_q[-\ln p(\tilde{\psi},\tilde{s})] - H[q(\tilde{\psi}\,|\,\tilde{\mu})]$$

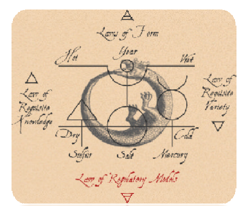
Maximum entropy principle

Ergodic theorem

$$\int dt F(t) \geq -\int dt \ln p(\tilde{s}(t)\,|\,m) = H[p(\tilde{s}\,|\,m)]$$

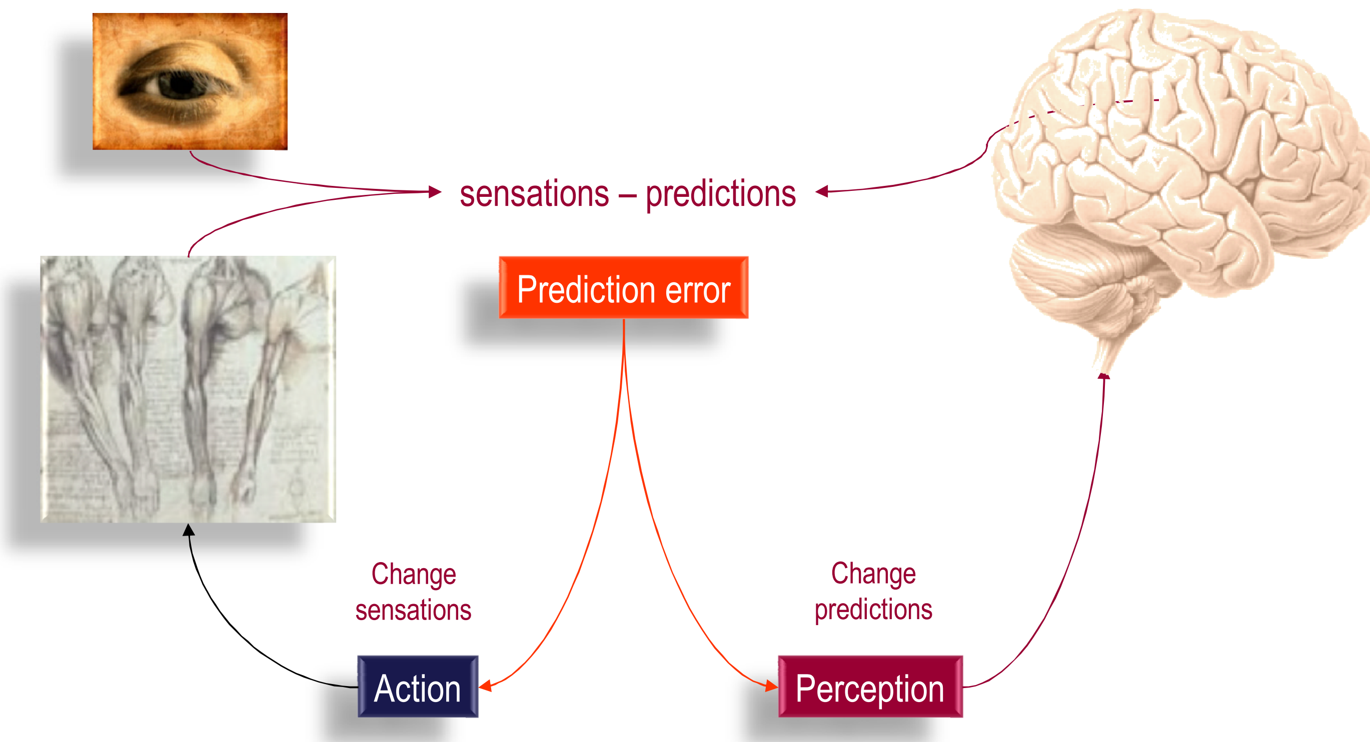
Minimum entropy principle



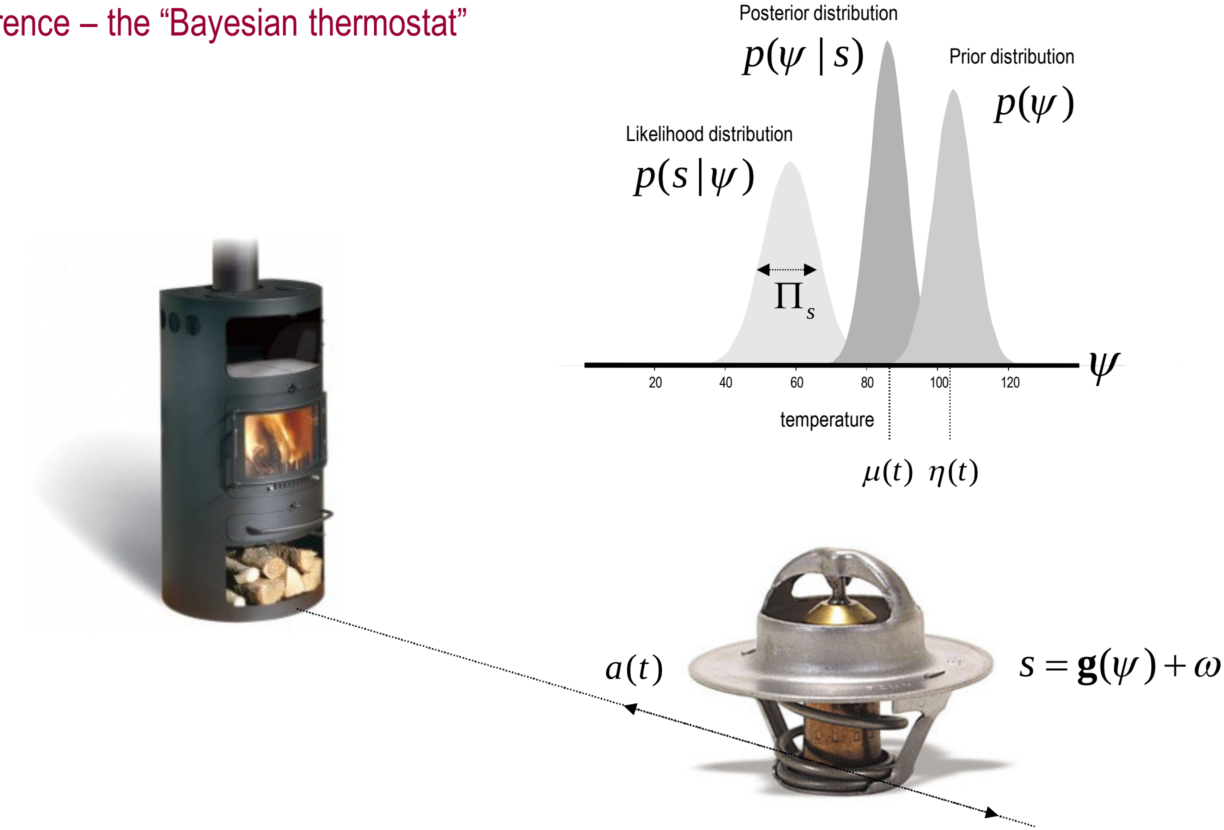
Self organisation and the principle of least action

How can we minimize surprise (prediction error)?



…action and perception minimise free energy

Action as inference – the "Bayesian thermostat"



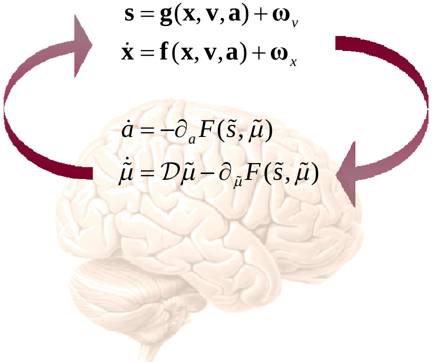
Perception $\mu = \arg\min_{\mu} F(s,\mu,\eta) = \arg\min_{\mu} \left\{ \Pi_s (s(a) - g(\mu))^2 + \Pi_\eta (\mu - \eta)^2 \right\}$

Action $a = \arg\min_{a} F(s,\mu,\eta) = \arg\min_{a} \left\{ \Pi_s (s(a) - g(\mu))^2 + \Pi_\eta (\mu - \eta)^2 \right\}$

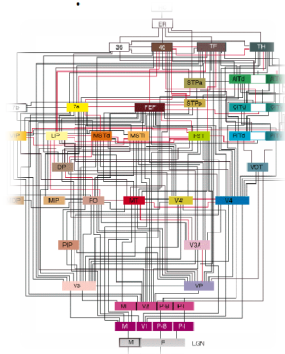
Free energy minimisation → Generative model → Predictive coding with reflexes

$$\mathbf{s} = \mathbf{g}(\mathbf{x},\mathbf{v},\mathbf{a}) + \boldsymbol{\omega}_v$$
$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x},\mathbf{v},\mathbf{a}) + \boldsymbol{\omega}_x$$

$$\dot{a} = -\partial_a F(\tilde{s},\tilde{\mu})$$
$$\dot{\tilde{\mu}} = \mathcal{D}\tilde{\mu} - \partial_{\tilde{\mu}} F(\tilde{s},\tilde{\mu})$$



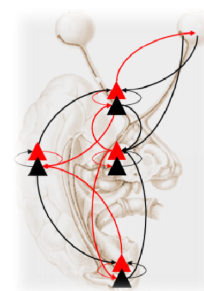
$$s = g^{(1)}(x^{(1)},v^{(1)}) + \omega_v^{(1)}$$
$$\dot{x}^{(1)} = f^{(1)}(x^{(1)},v^{(1)}) + \omega_x^{(1)}$$
$$\vdots$$
$$v^{(i-1)} = g^{(i)}(x^{(i)},v^{(i)}) + \omega_v^{(i)}$$
$$\dot{x}^{(i)} = f^{(i)}(x^{(i)},v^{(i)}) + \omega_x^{(i)}$$
$$\vdots$$



$$\xi_v^{(i)} = \Pi_v^{(i)}\tilde{\varepsilon}_v^{(i)} = \Pi_v^{(i)}(\tilde{\mu}_v^{(i-1)} - g^{(i)}(\tilde{\mu}_x^{(i)},\tilde{\mu}_v^{(i)}))$$
$$\xi_x^{(i)} = \Pi_x^{(i)}\tilde{\varepsilon}_x^{(i)} = \Pi_x^{(i)}(\mathcal{D}\tilde{\mu}_x^{(i)} - f^{(i)}(\tilde{\mu}_x^{(i)},\tilde{\mu}_v^{(i)}))$$

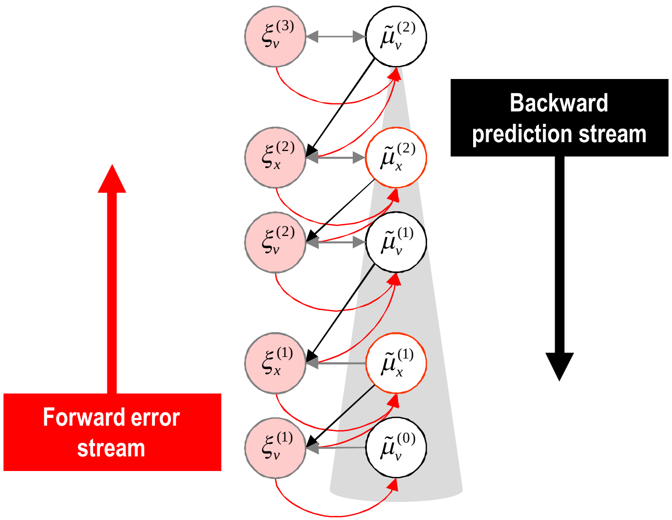
$$\dot{\tilde{\mu}}_v^{(i)} = D\tilde{\mu}_v^{(i)} - \partial_{\tilde{v}}\tilde{\varepsilon}^{(i)}\cdot\xi^{(i)} - \xi_v^{(i+1)}$$
$$\dot{\tilde{\mu}}_x^{(i)} = D\tilde{\mu}_x^{(i)} - \partial_{\tilde{x}}\tilde{\varepsilon}^{(i)}\cdot\xi^{(i)}$$

$$\dot{a} = -(\partial_a\tilde{\varepsilon}_v^{(1)})\cdot\xi_v^{(1)}$$



From models to perception

A simple hierarchy



$$pa(i) = i$$

$$\tilde{s} = \tilde{v}^{(0)}$$

Generative model

$$D\tilde{x}^{(i)} = f^{(i)}(\tilde{x}^{(i)}, \tilde{v}^{(i)}) + \tilde{\omega}_x^{(i)}$$

$$\tilde{v}^{(i-1)} = g^{(i)}(\tilde{x}^{(i)}, \tilde{v}^{(i)}) + \tilde{\omega}_v^{(i)}$$

Model inversion (inference)

Expectations:

$$\dot{\tilde{\mu}}_v^{(i)} = D\tilde{\mu}_v^{(i)} - \partial_{\tilde{v}}\tilde{\varepsilon}^{(i)} \cdot \xi^{(i)} - \xi_v^{(i+1)}$$

$$\dot{\tilde{\mu}}_x^{(i)} = D\tilde{\mu}_x^{(i)} - \partial_{\tilde{x}}\tilde{\varepsilon}^{(i)} \cdot \xi^{(i)}$$

Predictions:

$$g^{(i)} = g^{(i)}(\tilde{\mu}_x^{(i)}, \tilde{\mu}_v^{(i)})$$

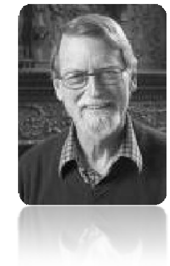
$$f^{(i)} = f^{(i)}(\tilde{\mu}_x^{(i)}, \tilde{\mu}_v^{(i)})$$

Prediction errors:

$$\xi_v^{(i)} = \Pi_v^{(i)}\tilde{\varepsilon}_v^{(i)} = \Pi_v^{(i)}(\tilde{\mu}_v^{(i-1)} - g^{(i)})$$

$$\xi_x^{(i)} = \Pi_x^{(i)}\tilde{\varepsilon}_x^{(i)} = \Pi_x^{(i)}(\mathcal{D}\tilde{\mu}_x^{(i)} - f^{(i)})$$

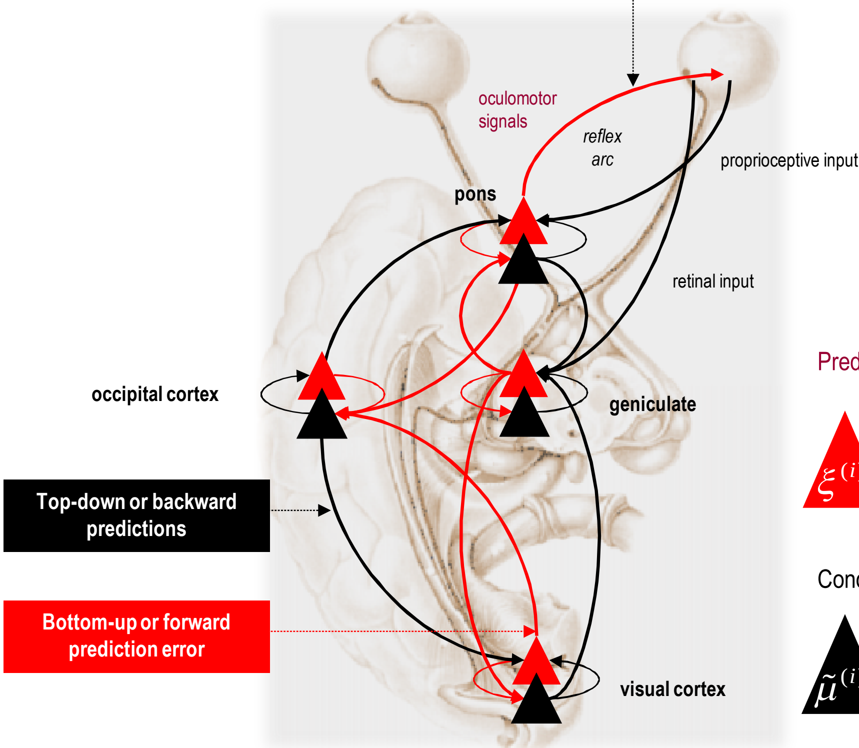
David Mumford



Predictive coding with reflexes

Action

$$\dot{a} = -\partial_a \tilde{s} \cdot \xi_v^{(1)}$$



Perception

Prediction error (superficial pyramidal cells)

$\xi^{(i)}$

$$\xi_v^{(i)} = \Pi_v^{(i)} \tilde{\varepsilon}_v^{(i)} = \Pi_v^{(i)}(\tilde{\mu}_v^{(i-1)} - g^{(i)}(\tilde{\mu}_x^{(i)}, \tilde{\mu}_v^{(i)}))$$

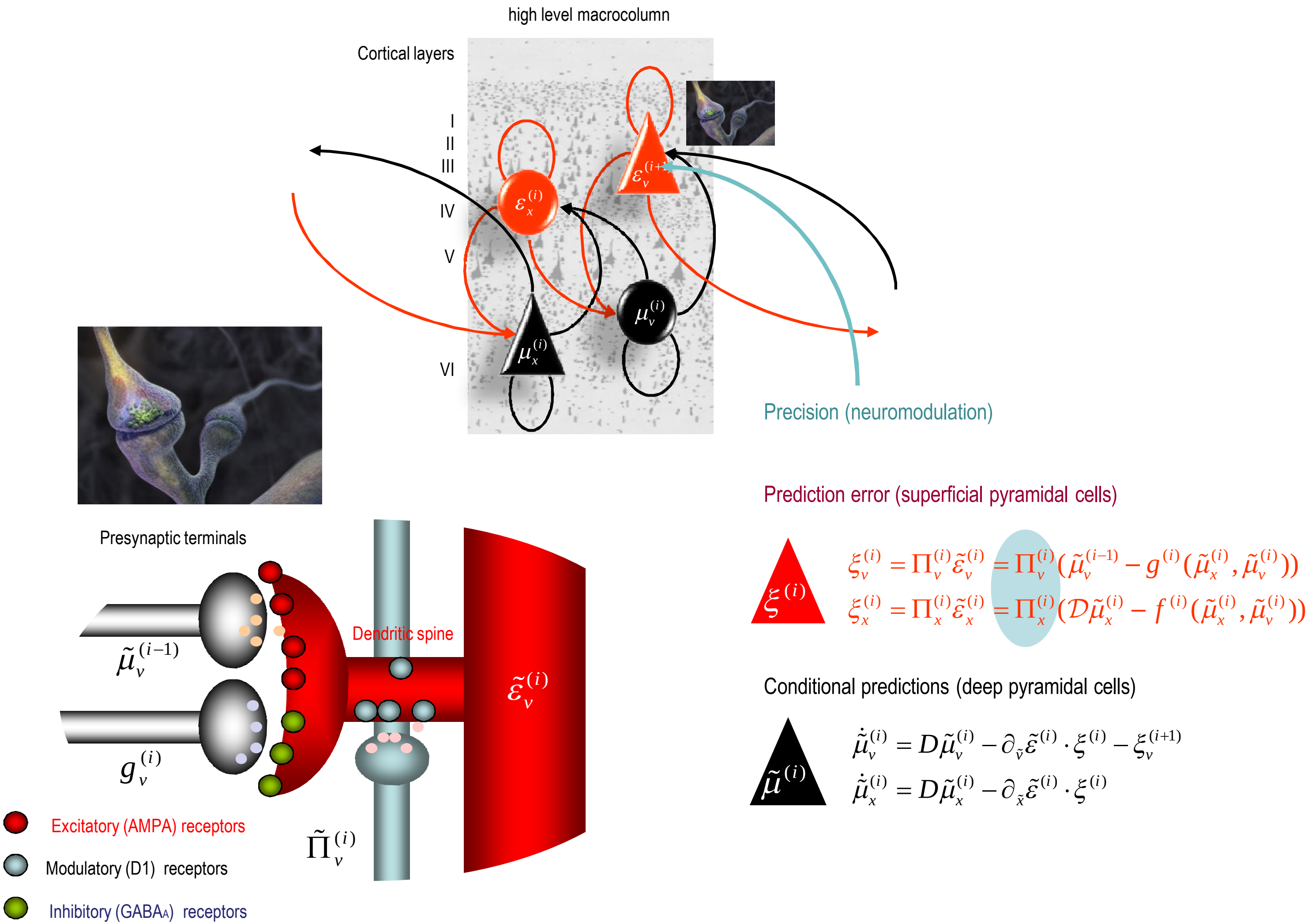
$$\xi_x^{(i)} = \Pi_x^{(i)} \tilde{\varepsilon}_x^{(i)} = \Pi_x^{(i)}(\mathcal{D}\tilde{\mu}_x^{(i)} - f^{(i)}(\tilde{\mu}_x^{(i)}, \tilde{\mu}_v^{(i)}))$$

Conditional predictions (deep pyramidal cells)

$\tilde{\mu}^{(i)}$

$$\dot{\tilde{\mu}}_v^{(i)} = D\tilde{\mu}_v^{(i)} - \partial_{\tilde{v}} \tilde{\varepsilon}^{(i)} \cdot \xi^{(i)} - \xi_v^{(i+1)}$$

$$\dot{\tilde{\mu}}_x^{(i)} = D\tilde{\mu}_x^{(i)} - \partial_{\tilde{x}} \tilde{\varepsilon}^{(i)} \cdot \xi^{(i)}$$



high level macrocolumn
Cortical layers
I
II
III
IV
V
VI
$\varepsilon_x^{(i)}$
$\varepsilon_v^{(i+1)}$
$\mu_v^{(i)}$
$\mu_x^{(i)}$
Precision (neuromodulation)
Prediction error (superficial pyramidal cells)
$\xi^{(i)}$
$\xi_v^{(i)} = \Pi_v^{(i)} \tilde{\varepsilon}_v^{(i)} = \Pi_v^{(i)}(\tilde{\mu}_v^{(i-1)} - g^{(i)}(\tilde{\mu}_x^{(i)}, \tilde{\mu}_v^{(i)}))$
$\xi_x^{(i)} = \Pi_x^{(i)} \tilde{\varepsilon}_x^{(i)} = \Pi_x^{(i)}(\mathcal{D}\tilde{\mu}_x^{(i)} - f^{(i)}(\tilde{\mu}_x^{(i)}, \tilde{\mu}_v^{(i)}))$
Conditional predictions (deep pyramidal cells)
$\tilde{\mu}^{(i)}$
$\dot{\tilde{\mu}}_v^{(i)} = D\tilde{\mu}_v^{(i)} - \partial_{\tilde{v}}\tilde{\varepsilon}^{(i)} \cdot \xi^{(i)} - \xi_v^{(i+1)}$
$\dot{\tilde{\mu}}_x^{(i)} = D\tilde{\mu}_x^{(i)} - \partial_{\tilde{x}}\tilde{\varepsilon}^{(i)} \cdot \xi^{(i)}$
Presynaptic terminals
Dendritic spine
$\tilde{\mu}_v^{(i-1)}$
$g_v^{(i)}$
$\tilde{\varepsilon}_v^{(i)}$
$\tilde{\Pi}_v^{(i)}$
Excitatory (AMPA) receptors
Modulatory (D1) receptors
Inhibitory (GABAA) receptors

Biological agents resist the second law of thermodynamics

They must minimize their average surprise (entropy)

They minimize surprise by suppressing prediction error (free-energy)

Prediction error can be reduced by changing predictions (perception)

Prediction error can be reduced by changing sensations (action)

Perception entails recurrent message passing in the brain to optimize predictions

Action makes predictions come true (and minimizes surprise)

Both action and perception depend the precision of prediction errors

Overview

The free-energy principle

- Action and perception
- Generative models
- Predictive coding

Perception

- Birdsong and attractors
- Deep models
- Simulated lesions and false inference

Action

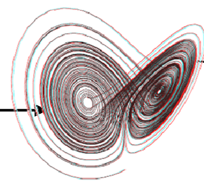
- Affordance and attractors
- Deep models
- Simulated lesions and false inference

Generating bird songs with attractors

$$v = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$$

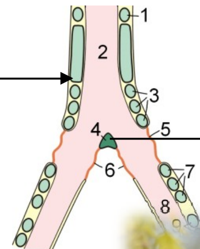
causal states

HVC

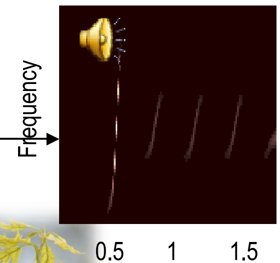


hidden states

Syrinx

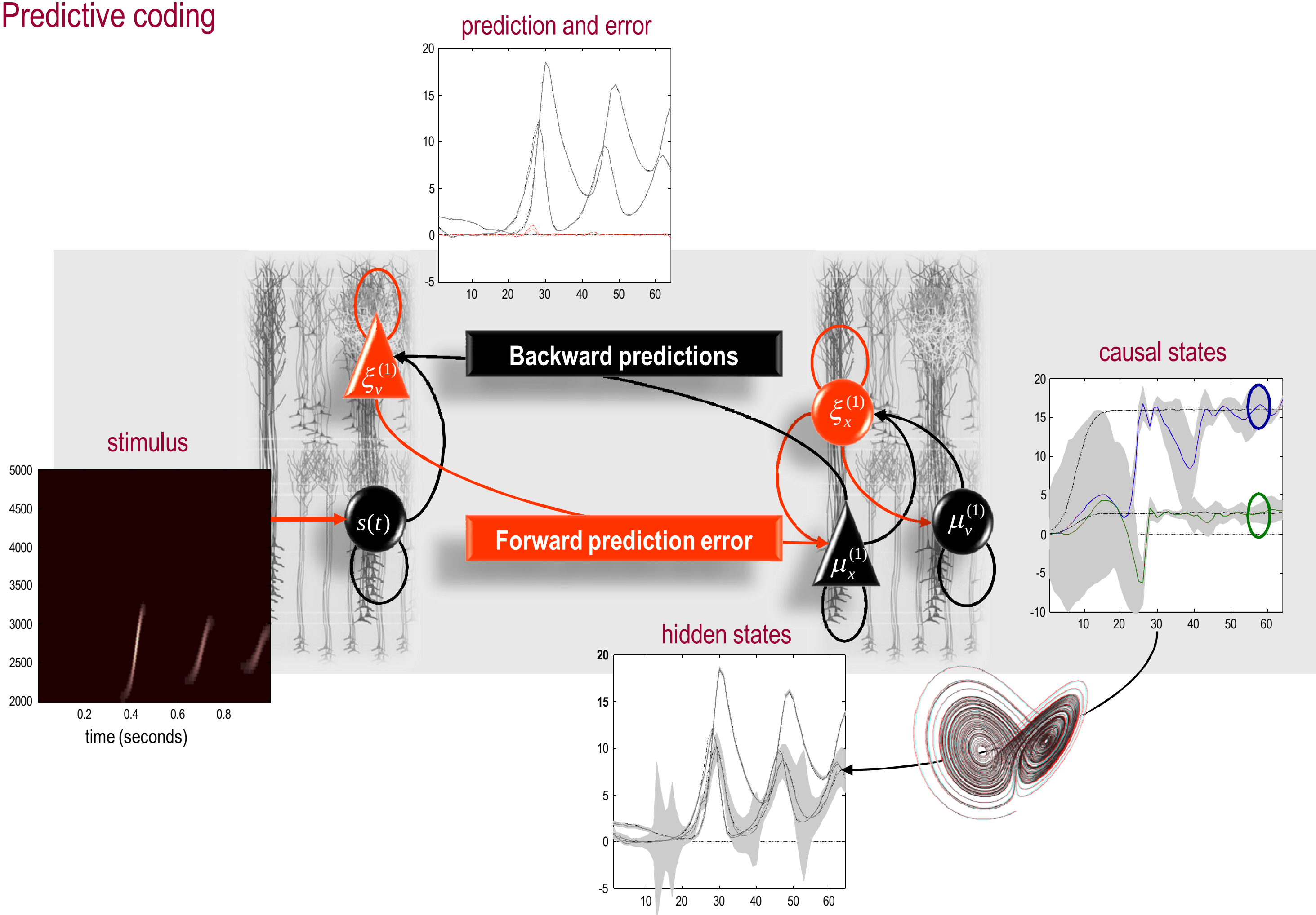


Sonogram

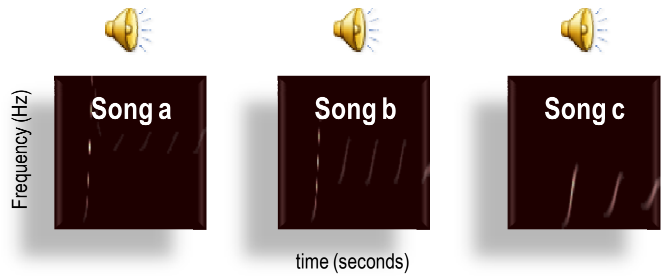


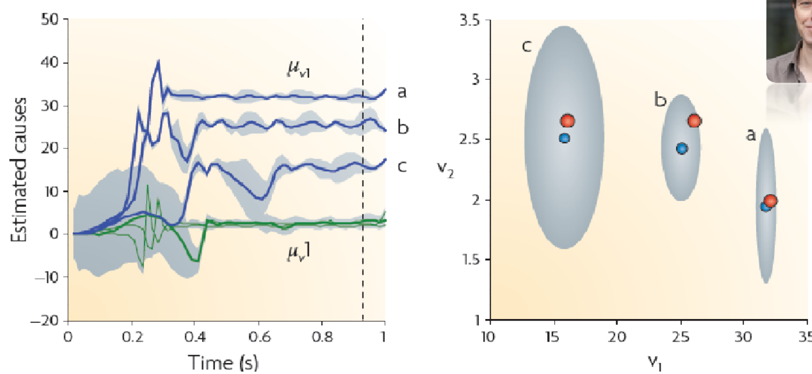
$$f^{(1)} = \begin{bmatrix} 18x_2^{(1)} - 18x_1^{(1)} \\ v_1^{(1)}x_1^{(1)} - 2x_3^{(1)}x_1^{(1)} - x_2^{(1)} \\ 2x_1^{(1)}x_2^{(1)} - v_2^{(1)}x_3^{(1)} \end{bmatrix}$$

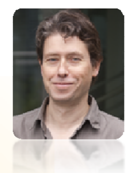
Predictive coding



Perceptual categorization

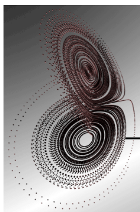




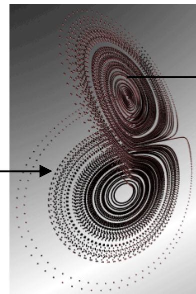


Hierarchical attractors: sequences of sequences

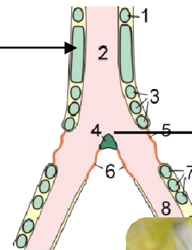
Neuronal hierarchy



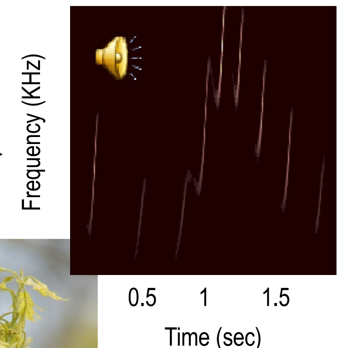
$v_1^{(1)}$
$v_2^{(1)}$



Syrinx

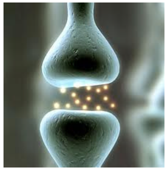


sonogram

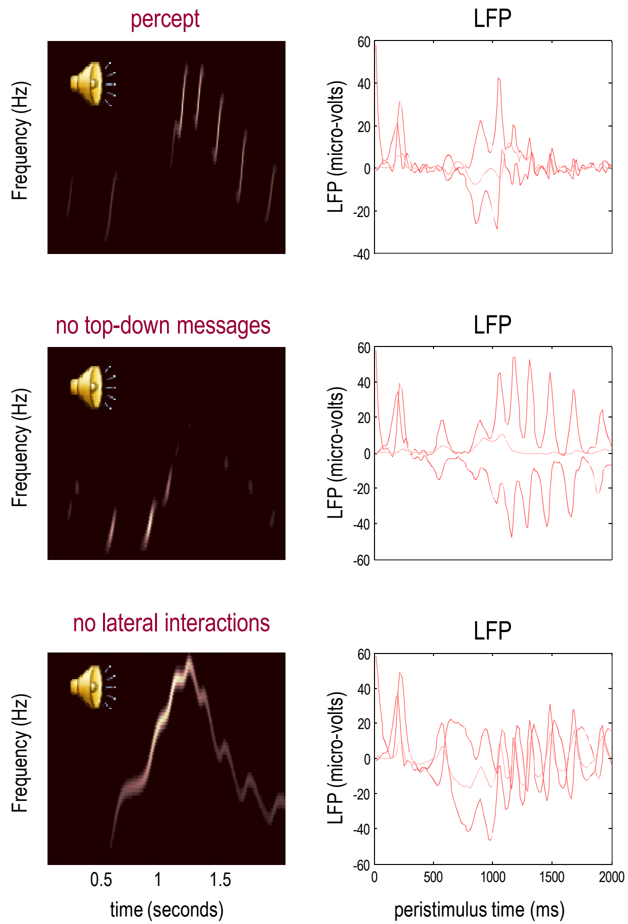


$$f^{(2)} = \begin{bmatrix} 18x_2^{(2)} - 18x_1^{(2)} \\ 32x_1^{(2)} - 2x_3^{(2)}x_1^{(2)} - x_2^{(2)} \\ 2x_1^{(2)}x_2^{(2)} - \frac{8}{3}x_3^{(2)} \end{bmatrix} \qquad f^{(1)} = \begin{bmatrix} 18x_2^{(1)} - 18x_1^{(1)} \\ v_1^{(1)}x_1^{(1)} - 2x_3^{(1)}x_1^{(1)} - x_2^{(1)} \\ 2x_1^{(1)}x_2^{(1)} - v_2^{(1)}x_3^{(1)} \end{bmatrix}$$

$$g^{(2)} = \begin{bmatrix} x_2^{(2)} \\ x_3^{(2)} \end{bmatrix} = \begin{bmatrix} v_1^{(1)} \\ v_2^{(1)} \end{bmatrix} \qquad g^{(1)} = \begin{bmatrix} x_2^{(1)} \\ x_3^{(1)} \end{bmatrix} = \begin{bmatrix} s_1 \\ s_2 \end{bmatrix}$$



Neuromodulatory lesions and false inference

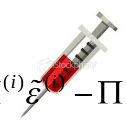


$$\dot{\tilde{\mu}}_v^{(i)} = \mathcal{D}\tilde{\mu}_v^{(i)} - \partial_{\tilde{v}}\tilde{\varepsilon}^{(i)} \cdot \Pi^{(i)}\tilde{\varepsilon}^{(i)} - \Pi_v^{(i+1)}\tilde{\varepsilon}_v^{(i+1)}$$
$$\dot{\tilde{\mu}}_x^{(i)} = \mathcal{D}\tilde{\mu}_x^{(i)} - \partial_{\tilde{x}}\tilde{\varepsilon}^{(i)} \cdot \Pi^{(i)}\tilde{\varepsilon}^{(i)}$$

no structural priors

$$\dot{\tilde{\mu}}_v^{(i)} = \mathcal{D}\tilde{\mu}_v^{(i)} - \partial_{\tilde{v}}\tilde{\varepsilon}^{(i)} \cdot \Pi^{(i)}\tilde{\varepsilon}^{(i)} - \Pi_v^{(i+1)}\tilde{\varepsilon}_v^{(i+1)}$$
$$\dot{\tilde{\mu}}_x^{(i)} = \mathcal{D}\tilde{\mu}_x^{(i)} - \partial_{\tilde{x}}\tilde{\varepsilon}^{(i)} \cdot \Pi^{(i)}\tilde{\varepsilon}^{(i)}$$

no dynamical priors



$$\dot{\tilde{\mu}}_v^{(i)} = \mathcal{D}\tilde{\mu}_v^{(i)} - \partial_{\tilde{v}}\tilde{\varepsilon}^{(i)} \cdot \Pi^{(i)}\tilde{\varepsilon}^{(i)} - \Pi_v^{(i+1)}\tilde{\varepsilon}_v^{(i+1)}$$
$$\dot{\tilde{\mu}}_x^{(i)} = \mathcal{D}\tilde{\mu}_x^{(i)} - \partial_{\tilde{x}}\tilde{\varepsilon}^{(i)} \cdot \Pi^{(i)}\tilde{\varepsilon}^{(i)}$$

Overview

The free-energy principle

Action and perception
Generative models
Predictive coding

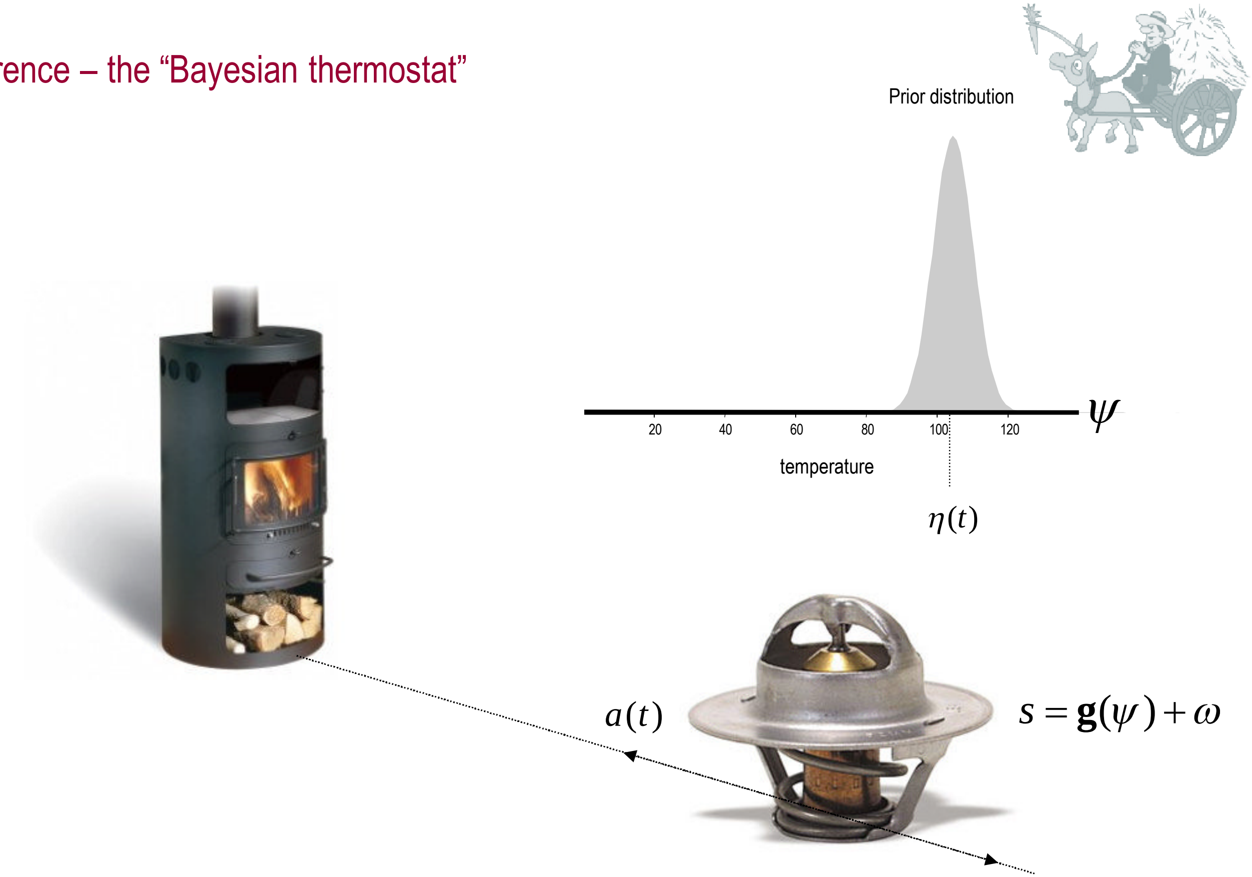
Perception

Birdsong and attractors
Deep models
Simulated lesions and false inference

Action

Affordance and attractors
Deep models
Simulated lesions and false inference

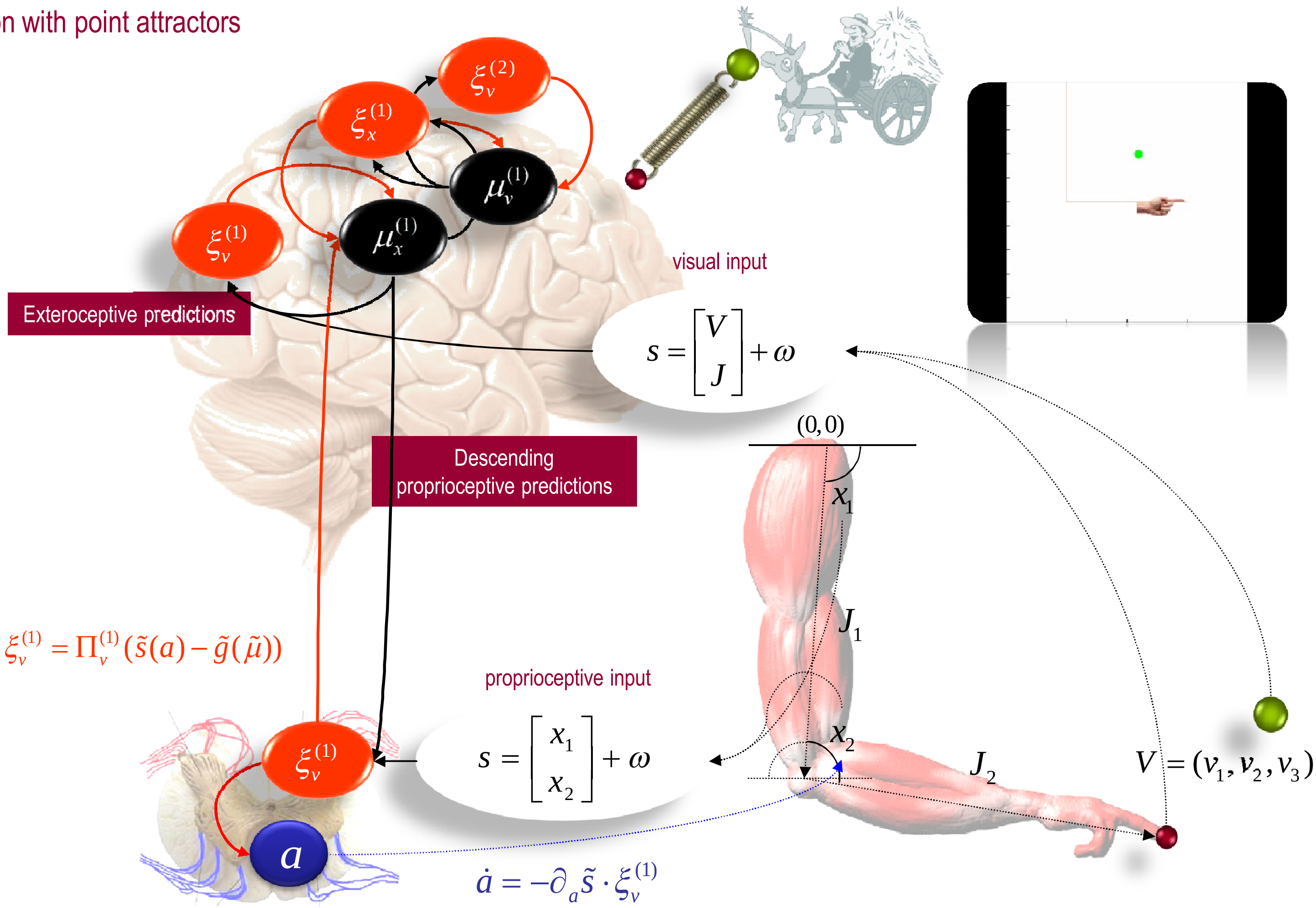
Action as inference – the “Bayesian thermostat”

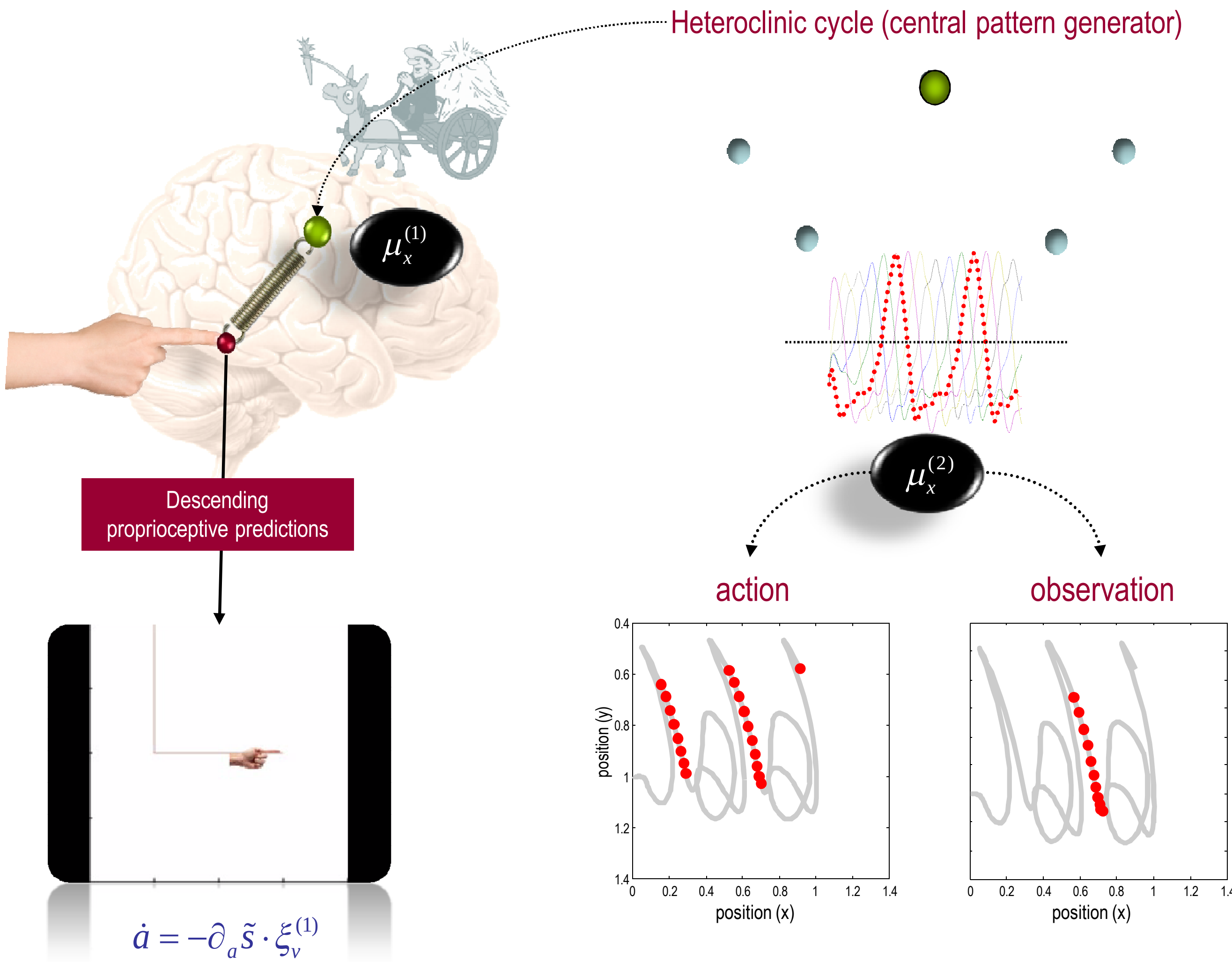


Perception $$\mu = \arg\min_{\mu} F(s,\mu,\eta) = \arg\min_{\mu} \left\{ \Pi_s (s(a) - g(\mu))^2 + \Pi_\eta (\mu - \eta)^2 \right\}$$

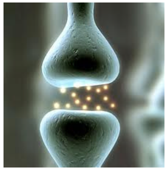
Action $$a = \arg\min_{a} F(s,\mu,\eta) = \arg\min_{a} \left\{ \Pi_s (s(a) - g(\mu))^2 + \Pi_\eta (\mu - \eta)^2 \right\}$$

Action with point attractors

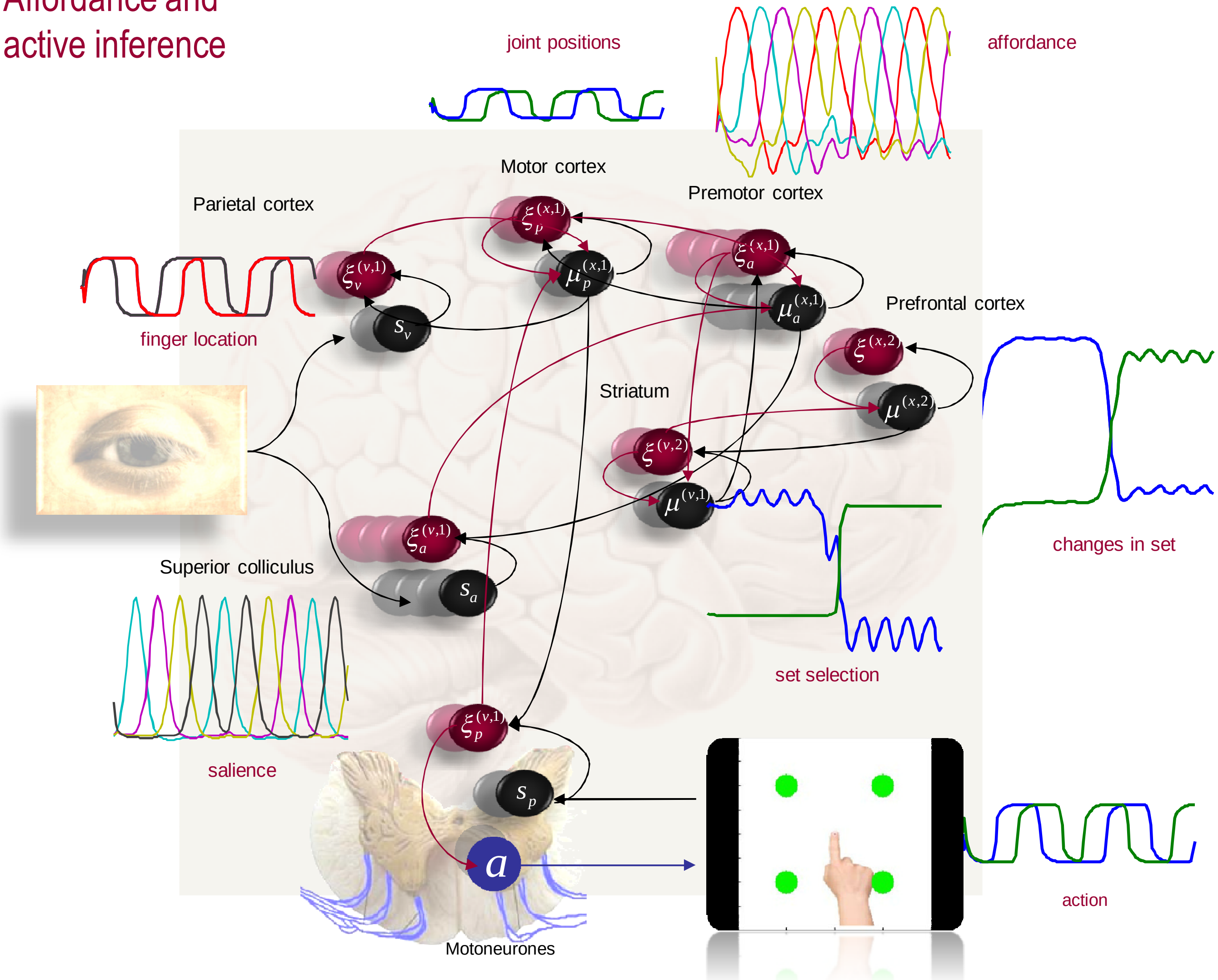


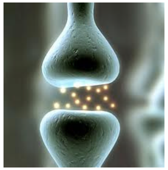


Heteroclinic cycle (central pattern generator)
$\mu_x^{(1)}$
Descending proprioceptive predictions
$\dot{a} = -\partial_a \tilde{s} \cdot \xi_v^{(1)}$
$\mu_x^{(2)}$
action
observation
position (y)
position (x)
position (x)



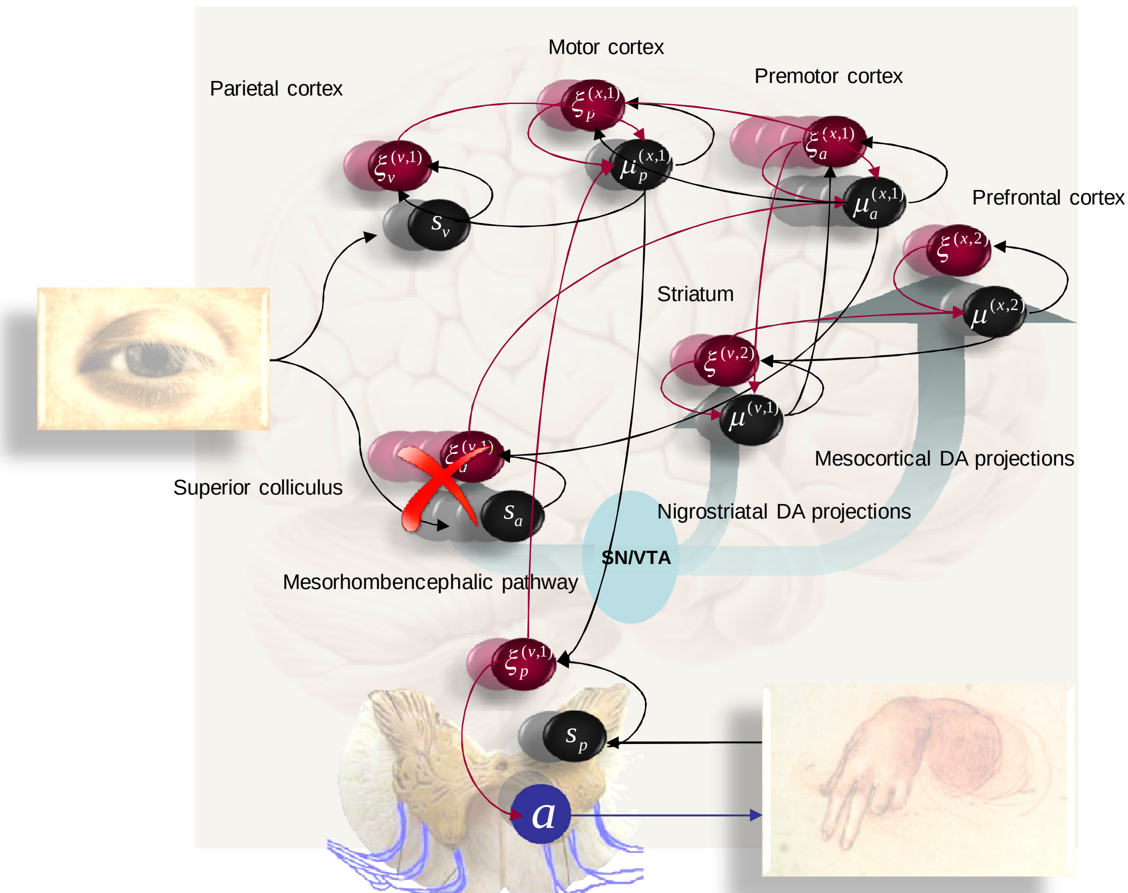
Affordance and active inference

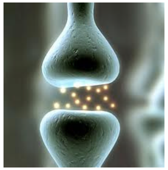




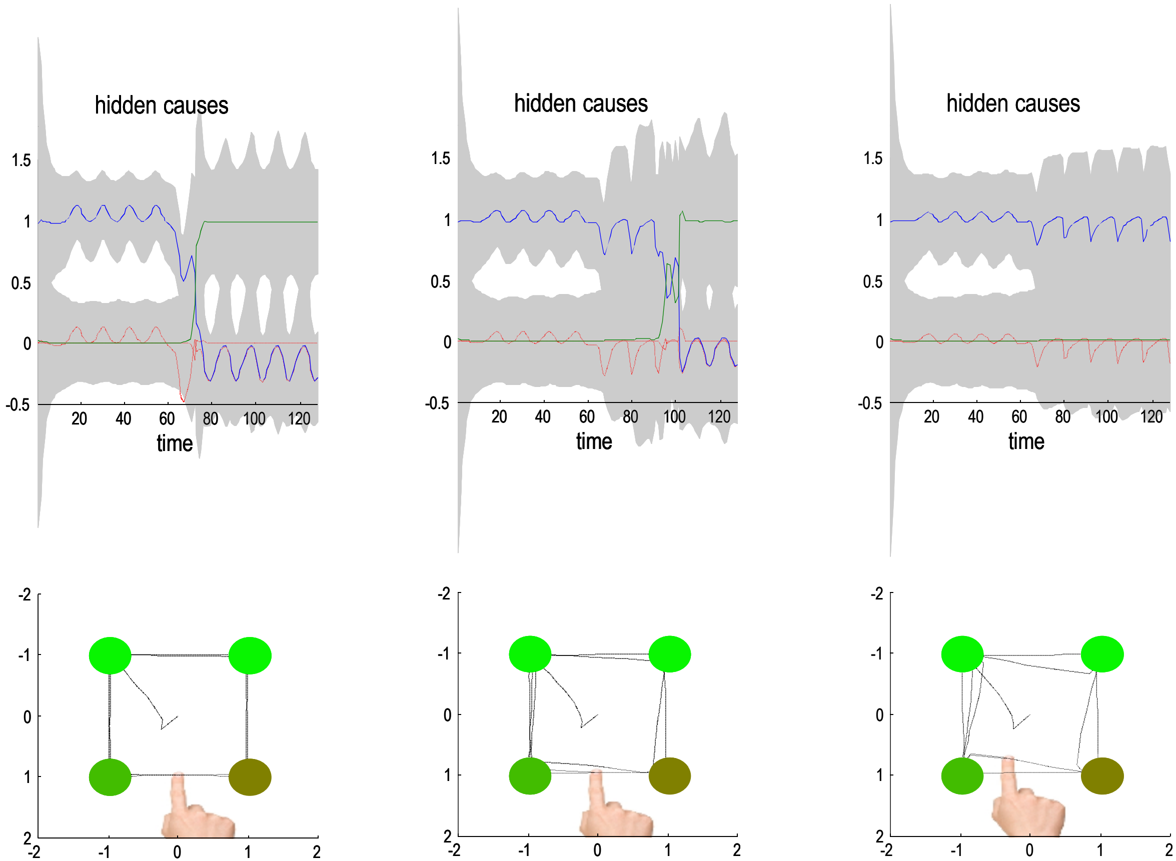
Dopamine and active inference

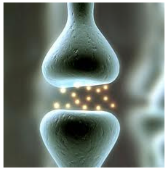
$$\xi^{(v,i)} = \Pi^{(v,i)}\tilde{\varepsilon}^{(v,i)} = \Pi^{(v,i)}(\tilde{\mu}^{(v,i-1)} - \tilde{g}^{(i)})$$
$$\xi^{(x,i)} = \Pi^{(x,i)}\tilde{\varepsilon}^{(x,i)} = \Pi^{(x,i)}(\mathcal{D}\tilde{\mu}^{(x,i)} - \tilde{f}^{(i)})$$



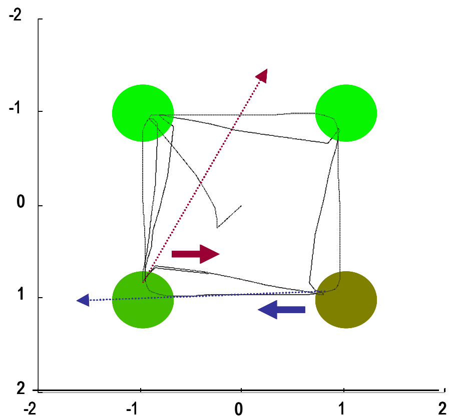


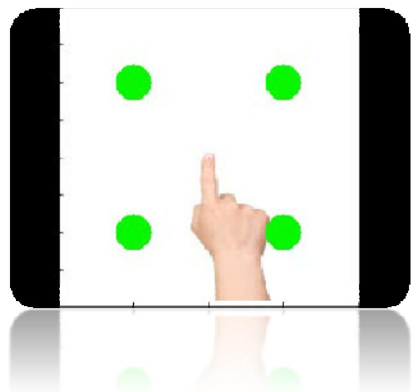
$$\tilde{\Pi}_a^{(1,v)} = \exp(5.0) \qquad \tilde{\Pi}_a^{(1,v)} = \exp(3.5) \qquad \tilde{\Pi}_a^{(1,v)} = \exp(2.5)$$

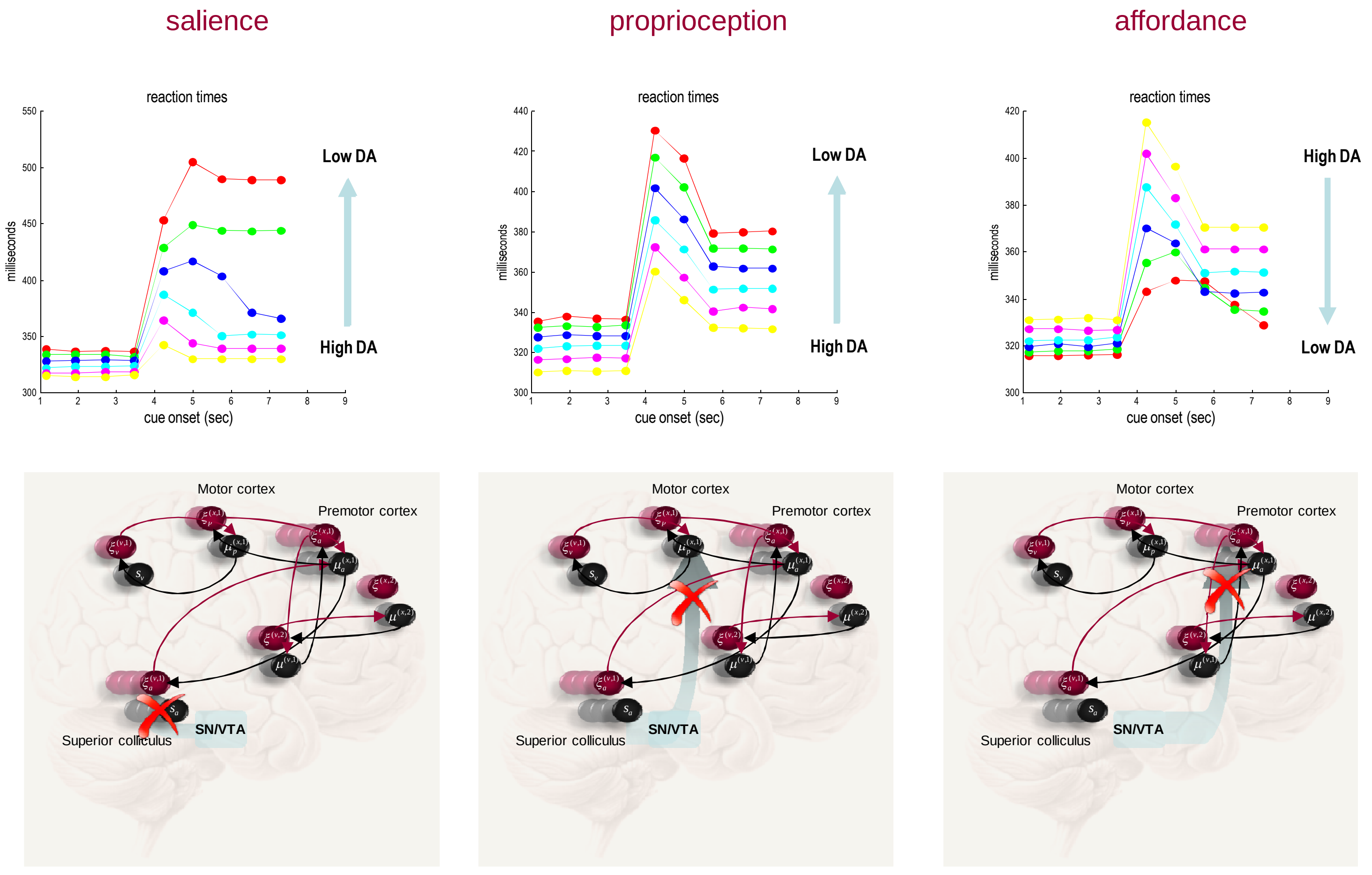




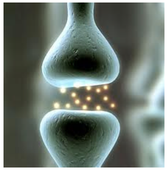
Uncertainty and perseveration



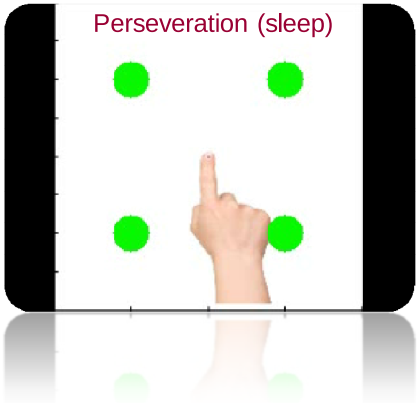


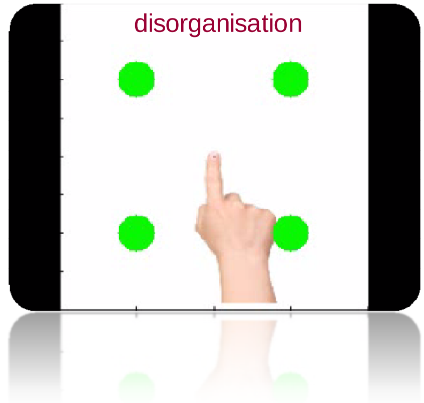


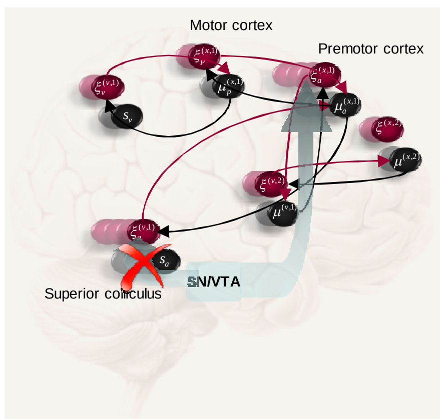
salience
proprioception
affordance
reaction times
milliseconds
cue onset (sec)
Low DA
High DA
High DA
Low DA
Motor cortex
Premotor cortex
Superior colliculus
SN/VTA

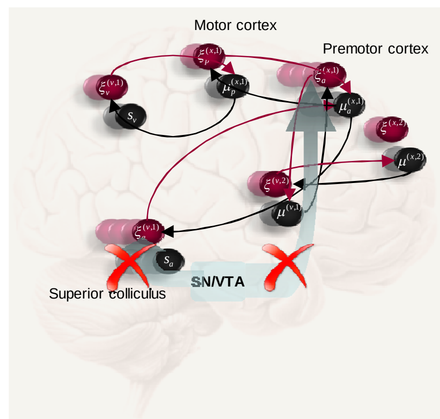


Uncertainty, perseveration and disorganisation









Thank you

And thanks to collaborators:

Rick Adams
Andre Bastos
Sven Bestmann
Jean Daunizeau
Harriet Brown
Lee Harrison
Tom FitzGerald
Stefan Kiebel
James Kilner
Jérémie Mattout
Read Montague
Rosalyn Moran
Will Penny
Klaas Stephan

And colleagues:

Andy Clark
Peter Dayan
Jörn Diedrichsen
Ray Dolan
Paul Fletcher
Pascal Fries
Geoffrey Hinton
James Hopkins
Jakob Hohwy
Henry Kennedy
Paul Verschure
Florentin Wörgötter

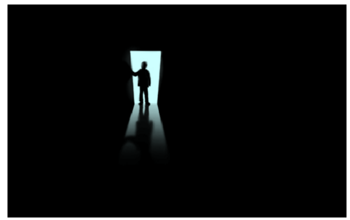
And many others

frontiers in PSYCHOLOGY
HYPOTHESIS AND THEORY ARTICLE
published: 08 May 2012
doi: 10.3389/fpsyg.2012.00130

Free-energy minimization and the dark-room problem

Karl Friston[1], Christopher Thornton[2] and Andy Clark[3]*

[1] *The Wellcome Trust Centre for Neuroimaging, University College London, London, UK*
[2] *Informatics, University of Sussex, Brighton, UK*
[3] *School of Philosophy, Psychology, and Language Sciences, University of Edinburgh, Edinburgh, UK*



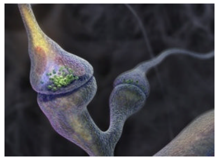
Searching to test hypotheses – life as an efficient experiment

$$
\begin{aligned}
H(S,\Psi) &= H(S \mid m) + H(\Psi \mid S) \\
&= \underbrace{E_t[-\ln p(\tilde{s}(t) \mid m)]}_{\text{Free energy principle}} + \underbrace{E_t[H(\Psi \mid S = \tilde{s}(t))]}_{\text{minimise uncertainty}}
\end{aligned}
$$

$$
\tilde{\eta}(t) = \arg\min_{\tilde{\eta}} \{H[q(\tilde{\psi} \mid \tilde{\mu}, \tilde{\eta})]\}
$$

Time-scale

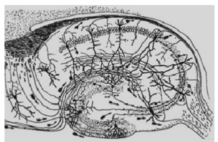
Free-energy minimisation leading to…



$10^{-3}\,s$

$10^{0}\,s$

Perception and Action: The optimisation of neuronal and neuromuscular activity to suppress prediction errors (or free-energy) based on generative models of sensory data.



$10^{3}\,s$

Learning and attention: The optimisation of synaptic gain and efficacy over seconds to hours, to encode the precisions of prediction errors and causal structure in the sensorium. This entails suppression of free-energy over time.

$10^{6}\,s$

Neurodevelopment: Model optimisation through activity-dependent pruning and maintenance of neuronal connections that are specified epigenetically

$10^{15}\,s$

Evolution: Optimisation of the average free-energy (free-fitness) over time and individuals of a given class (e.g., conspecifics) by selective pressure on the epigenetic specification of their generative models.