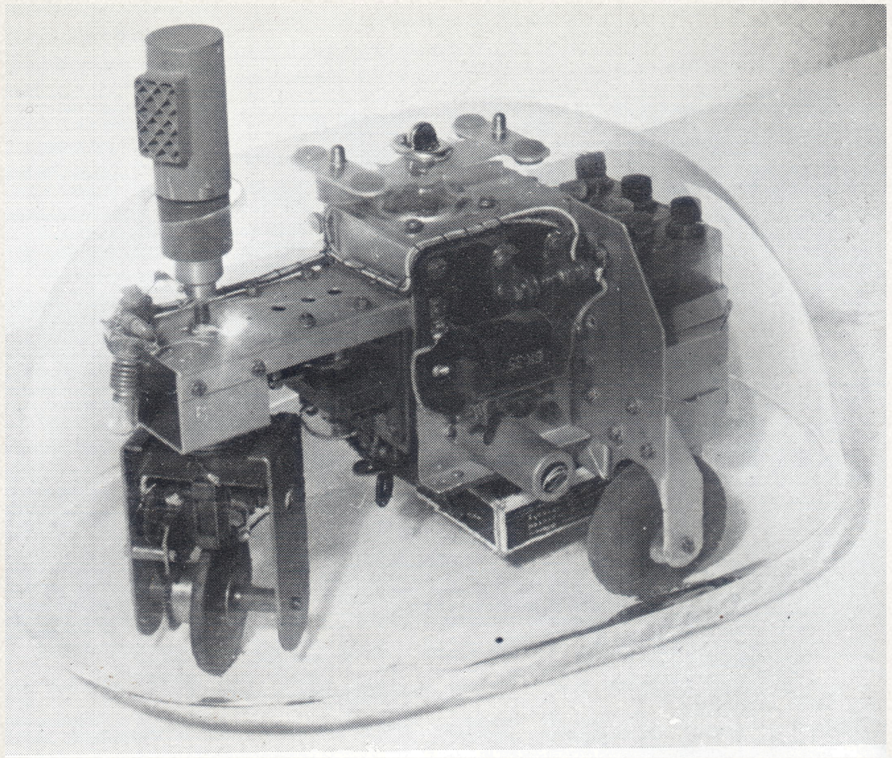
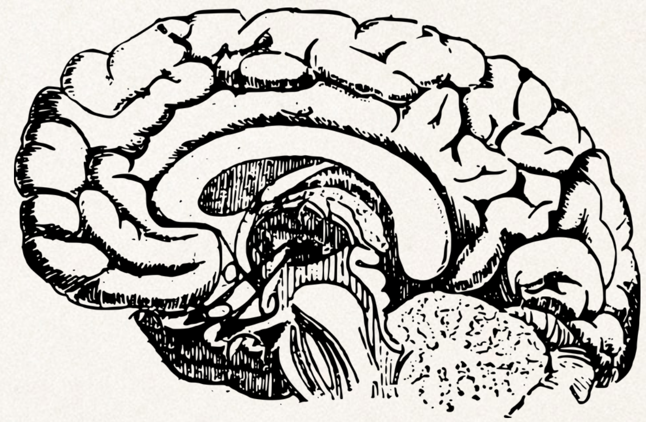




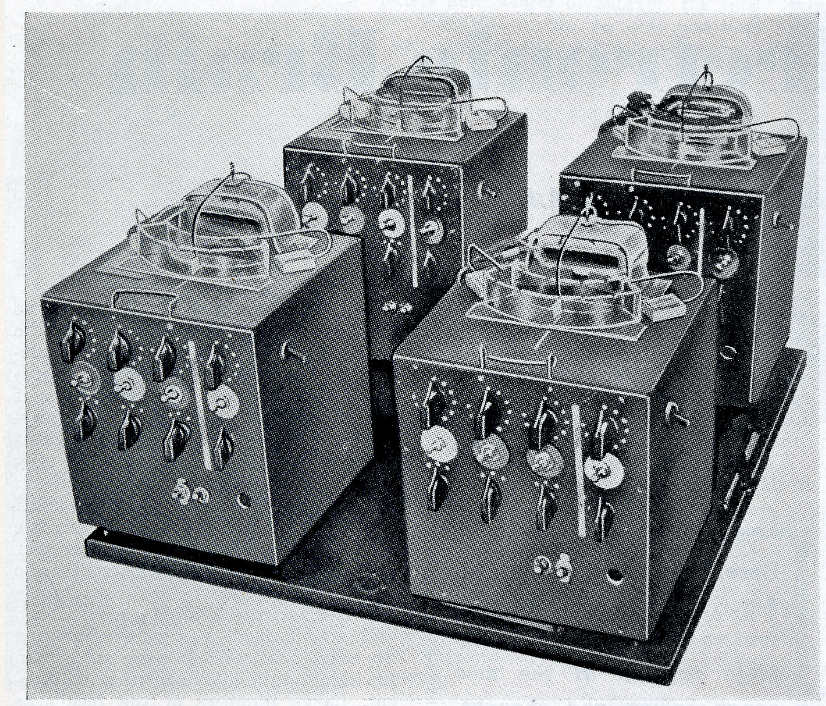


Fig. I—The homeostat, with its four units, each one of which reacts on all the others.

Nonmodular Architectures of Cognitive Systems
based on Active Inference

Manuel Baltieri

- EASy - CCNR - Sussex Neuroscience, University of Sussex, Brighton (UK)
- Visitor @ EON (ELSI Origins of life Network), ELSI, Tokyo Tech, Tokyo (Japan)

- Twitter: @manuelbaltieri

University of Sussex

10/07/2019 - Brighton

Roadmap

- Modularity, especially in cognitive science
- Separation principle in control theory (and neuroscience)
- An alternative in *active inference*

Modularity?

Intuitive definition

Decomposability (and near-decomposability) - Herbert A. Simon

Structural and functional - Watson and Pollack (2005)

“The modular mind” - Jerry A. Fodor

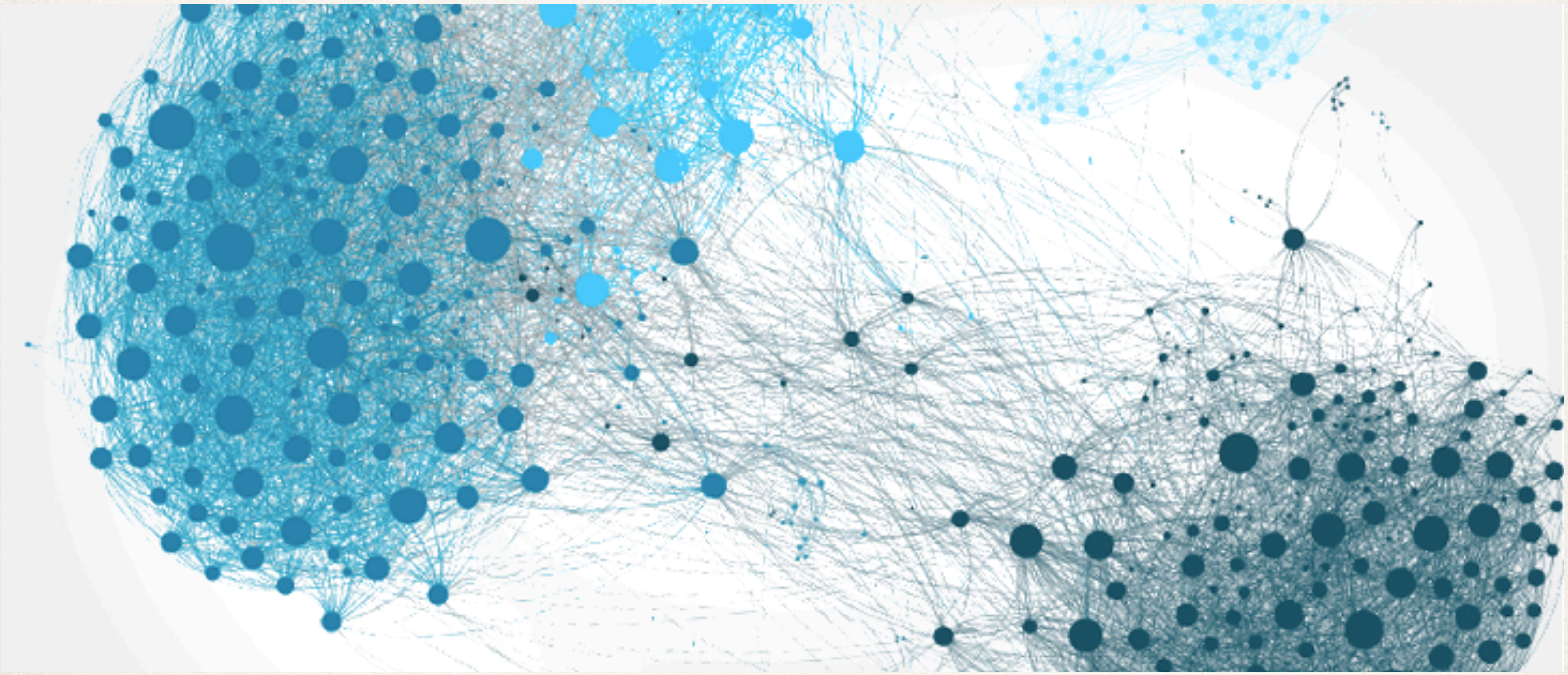
Modularity?

Intuitive definition

Decomposability (and near-decomposability) - Herbert A. Simon

Structural and functional - Watson and Pollack (2005)

“The modular mind” - Jerry A. Fodor



Modularity?

Intuitive definition

Decomposability (and near-decomposability) - Herbert A. Simon

Structural and functional - Watson and Pollack (2005)

“The modular mind” - Jerry A. Fodor

“The main theoretical findings from the approach can be summed up in two propositions:

(a) in a nearly decomposable system, the short run behavior of each of the component subsystems is approximately independent of the short-run behavior of the other components;

(b) in the long run, the behavior of any one of the components depends in only an aggregate way on the behavior of the other components. ”

– “The Architecture of Complexity”. Herbert A. Simon

(on near-decomposability)

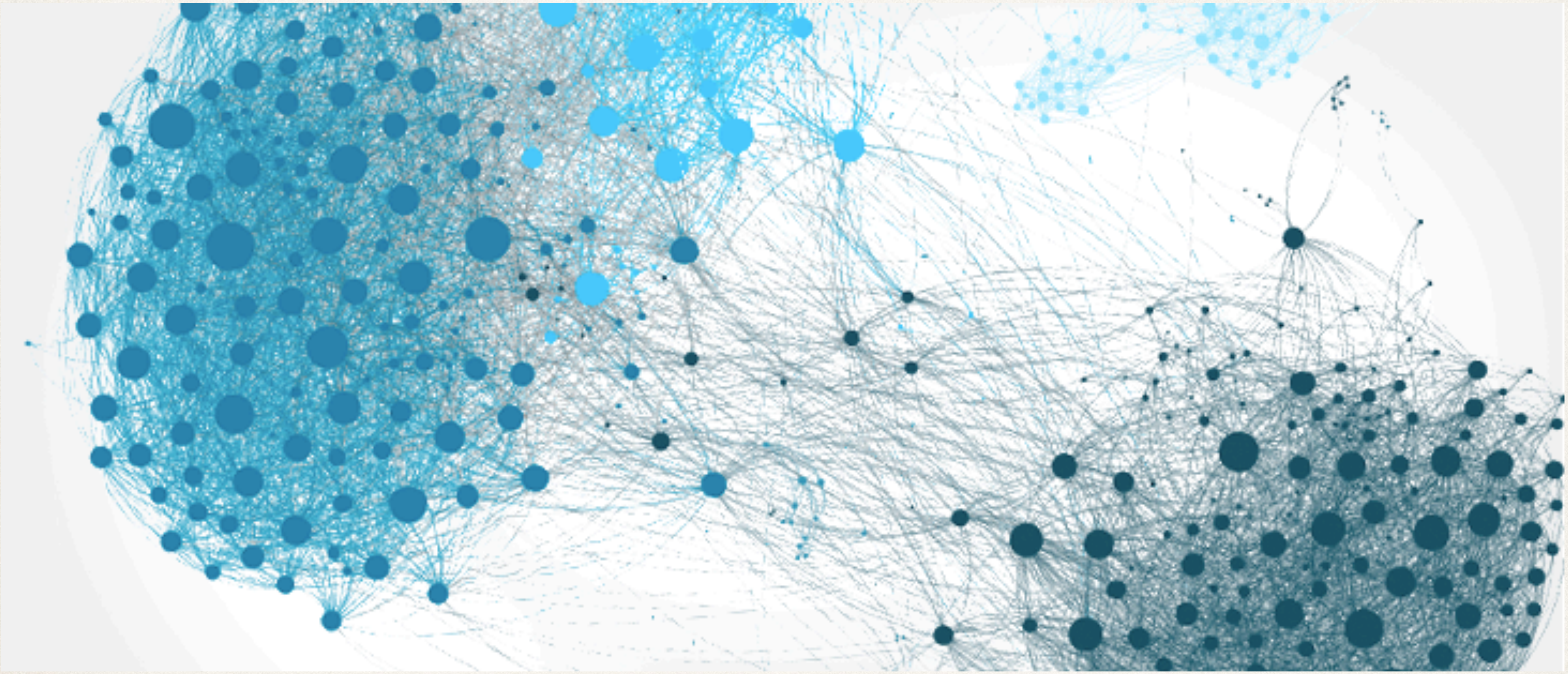
Modularity?

Intuitive definition

Decomposability (and near-decomposability) - Herbert A. Simon

Structural and functional - Watson and Pollack (2005)

“The modular mind” - Jerry A. Fodor



Modularity?

Intuitive definition

Decomposability (and near-decomposability) - Herbert A. Simon

Structural and functional - Watson and Pollack (2005)

“The modular mind” - Jerry A. Fodor

Modularity?

Intuitive definition

Decomposability (and near-decomposability) - Herbert A. Simon

Structural and functional - Watson and Pollack (2005)

"The modular mind" - Jerry A. Fodor

Fodorian modularity

Modules are:

- domain specific
- innately specified
- informationally encapsulated
- fast
- hardwired (neurally specific)
- autonomous
- not assembled

Fodorian modularity

Modules are:

- **domain specific**
- innately specified
- **informationally encapsulated**
- fast
- hardwired (neurally specific)
- **autonomous**
- not assembled

The sandwich of cognitive science

Perception

↓

Cognition

↓

Action

Cognitivism/ computationalism

Themes:

Feedforward computation, (informationally) independent components, symbol manipulation, central processing, input-output structure, information processing and encoding, (internal) representations, etc.

But the brain *is not* (just) a feedforward network of bottom-up signals

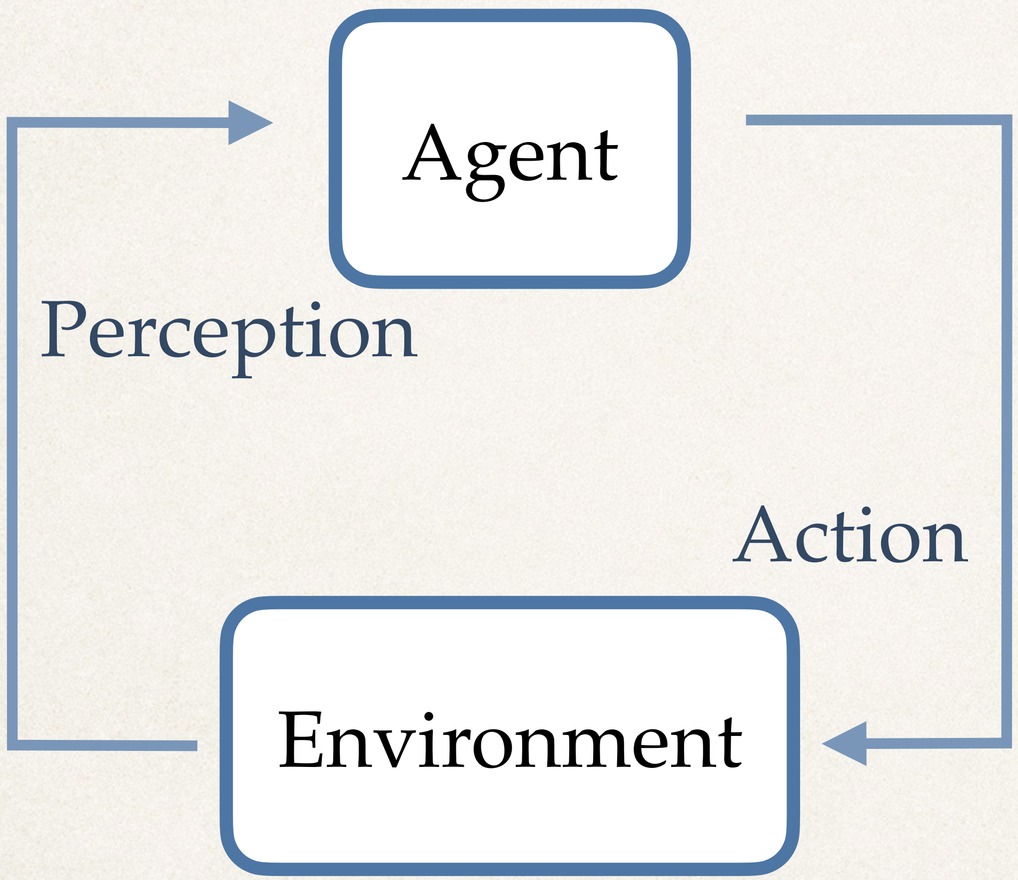
Feedback (top-down and lateral) connections are everywhere

The brain *is not* an isolated system

The brain resides in embodied and situated agents

Buckley, Christopher L., and Taro Toyoizumi. "A theory of how active behavior stabilises neural activity: Neural gain modulation by closed-loop environmental feedback." *PLoS computational biology* 14.1 (2018): e1005926.

The 4E views, agents and environments (and to an extent, RL)



4Es: Embodied, Enactive, Extended and Embedded

Themes:

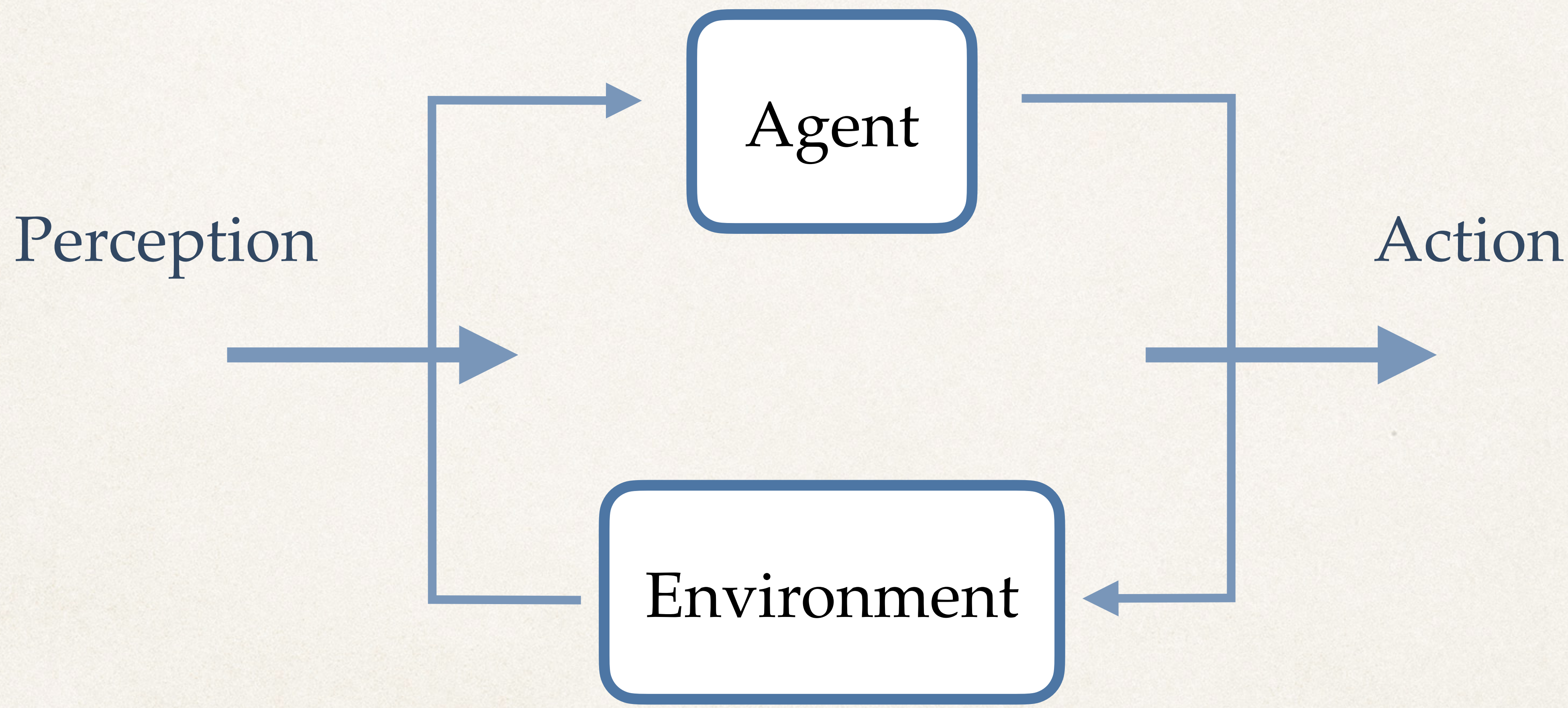
Feedback from environment, role of the body, situatedness, dynamics (time), autonomy, affordances, sensorimotor coupling, etc.

Examples:

- Cybernetics (Wiener, Ashby, …),
- Behaviour-based robotics (Brooks),
- Evolutionary robotics (Harvey, Husbands, Floreano, Bongard, …),
- Developmental robotics (Lungarella, …),
- Dynamical systems approaches (Beer),
- etc.

We like modules…

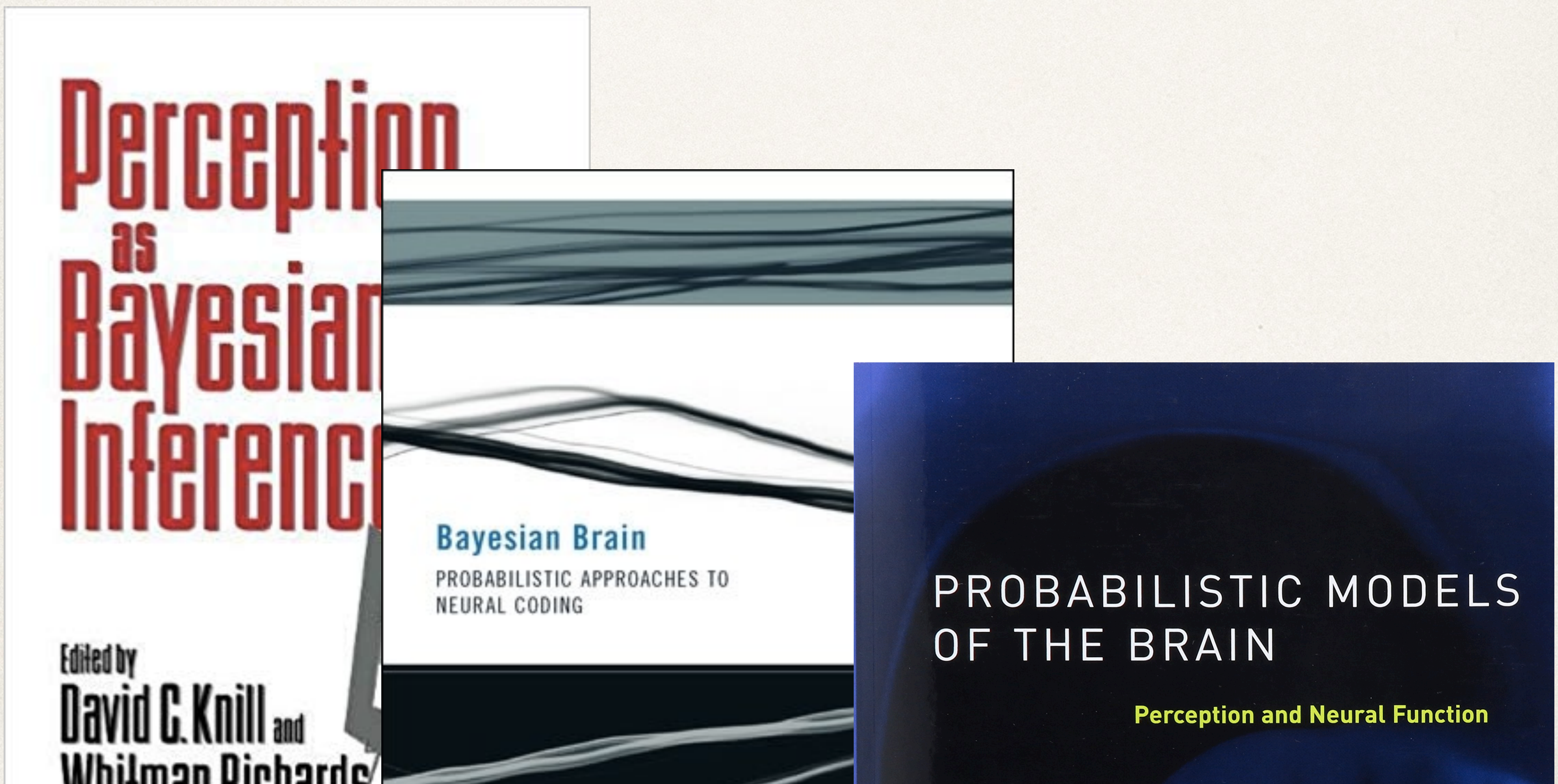
Input-output architectures (classical sandwich)



Meanwhile, in modern neuroscience…

Claim 1

Perception can be described as a process of (Bayesian) inference or estimation



Claim 2

Action can be described as a process of (optimal) control

718

Internal models for motor control and trajectory planning
Mitsuo Kawato

A number of internal model concepts are nc
neuroscience and cognitive science. These
supported by behavioral, neurophysiologica
furthermore, these models have had their st
functions revealed by such data. In particula
on inverse dynamics model learning is direc
unit recordings from cerebellar Purkinje cell
forward inverse models describing how dive
environments can be controlled and learnec
recently been proposed. The 'minimum varia
another major recent advance in the compu
motor control. This model integrates two fu
approaches on trajectory planning, strongly
both kinematic and dynamic internal models
movement planning and control.

Addresses
ATR Human Information Processing Research La

review

Computational principles of movement neuroscience

Daniel M. Wolpert[1] and Zoubin Ghahra

[1] Sobell Department of Neurophysiology, Institute of Neurolog
[2] Gatsby Computational Neuroscience Unit, Queen Square, U
Correspondence should be addressed to D.M.W. (wolpert@he

**Unifying principles of movement have em
review several of these principles and sho
control, estimation, prediction and learnin
ing from the computational approach pro**

The computational study of motor control is fund
cerned with the relationship between sensory sig
commands. The transformation from motor com
sensory consequences is governed by the physics

http://neurosci.nature.com

REVIEW

nature neuroscience

Optimality principles in sensorimotor control

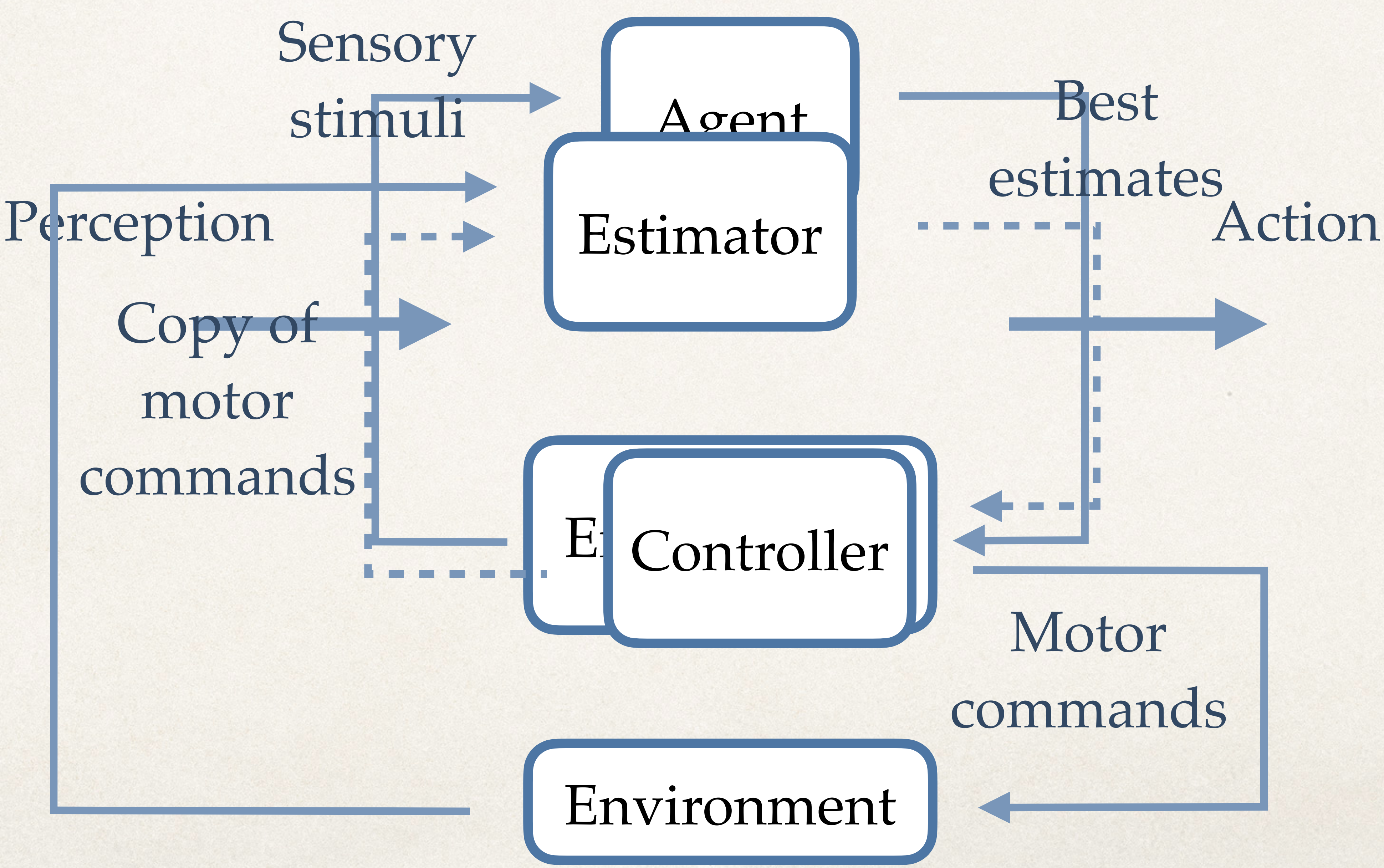
Emanuel Todorov

/natureneuroscience

We really like modules…

Perception and action combined in cognitive (neuro)science

Input-output architectures (classical view)



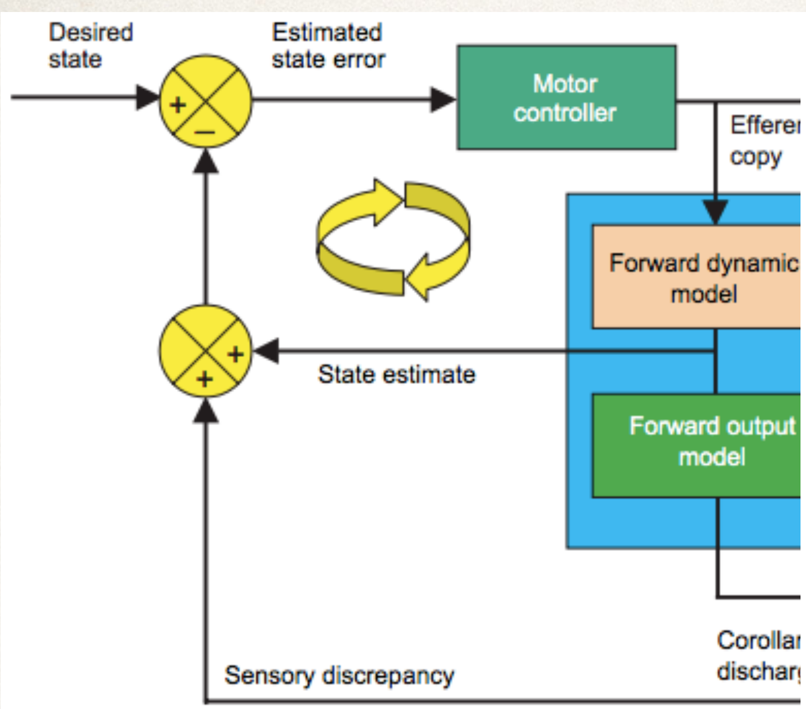
Claim 3 (often implicit)

The separation principle

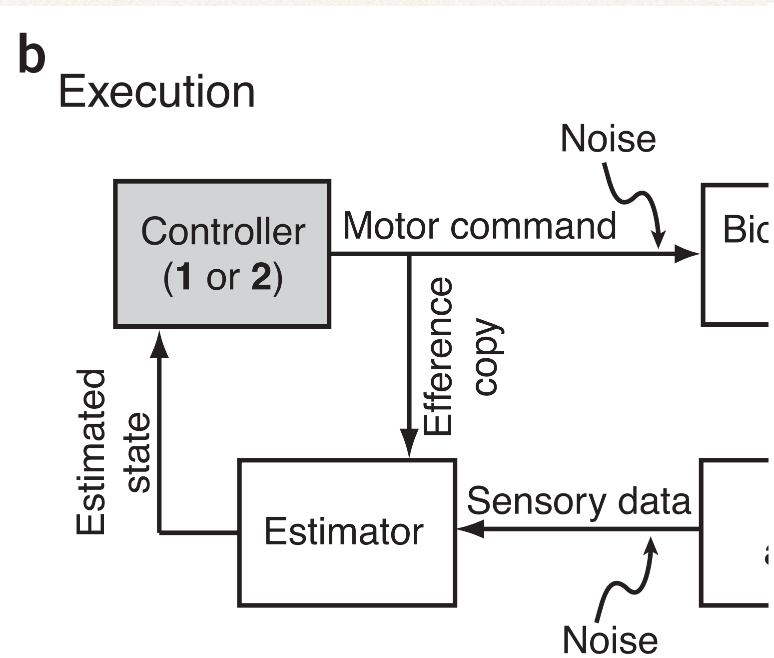
Classic result in control theory (cf. "certainty equivalence" in econometrics and separation principle in information theory) for **linear systems**:

LQG (Linear Quadratic Gaussian) control =

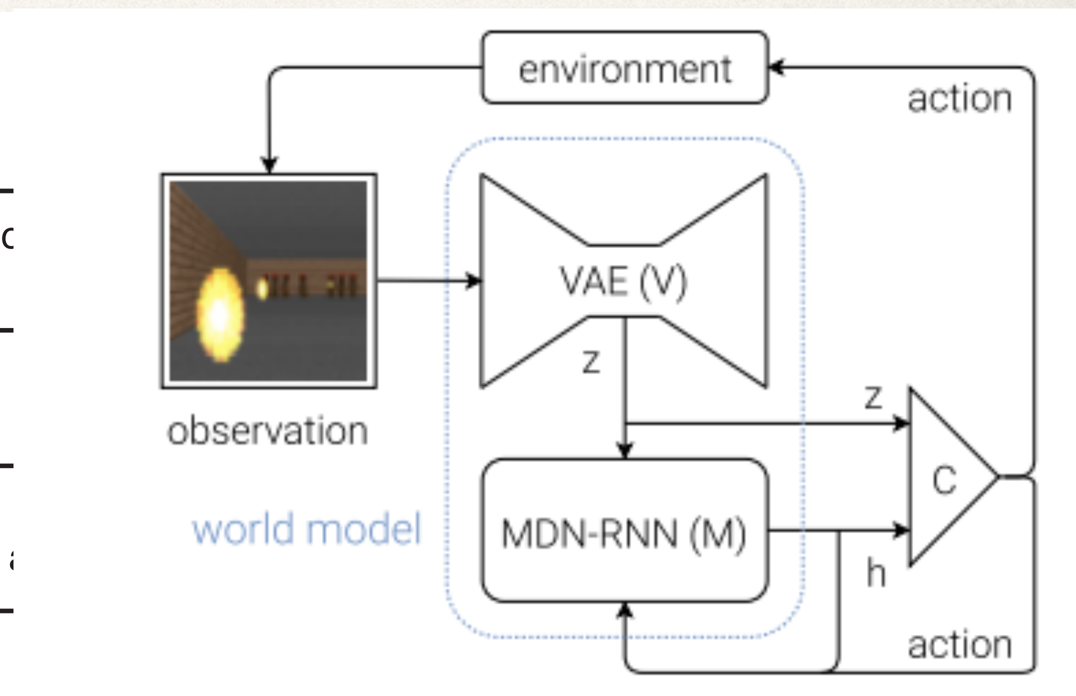
Kalman filter (**estimator**) + Linear quadratic regulator (**controller**)



Wolpert et al. (1998)



Todorov (2004)



Ha et al. (2018), but this one is "ok"

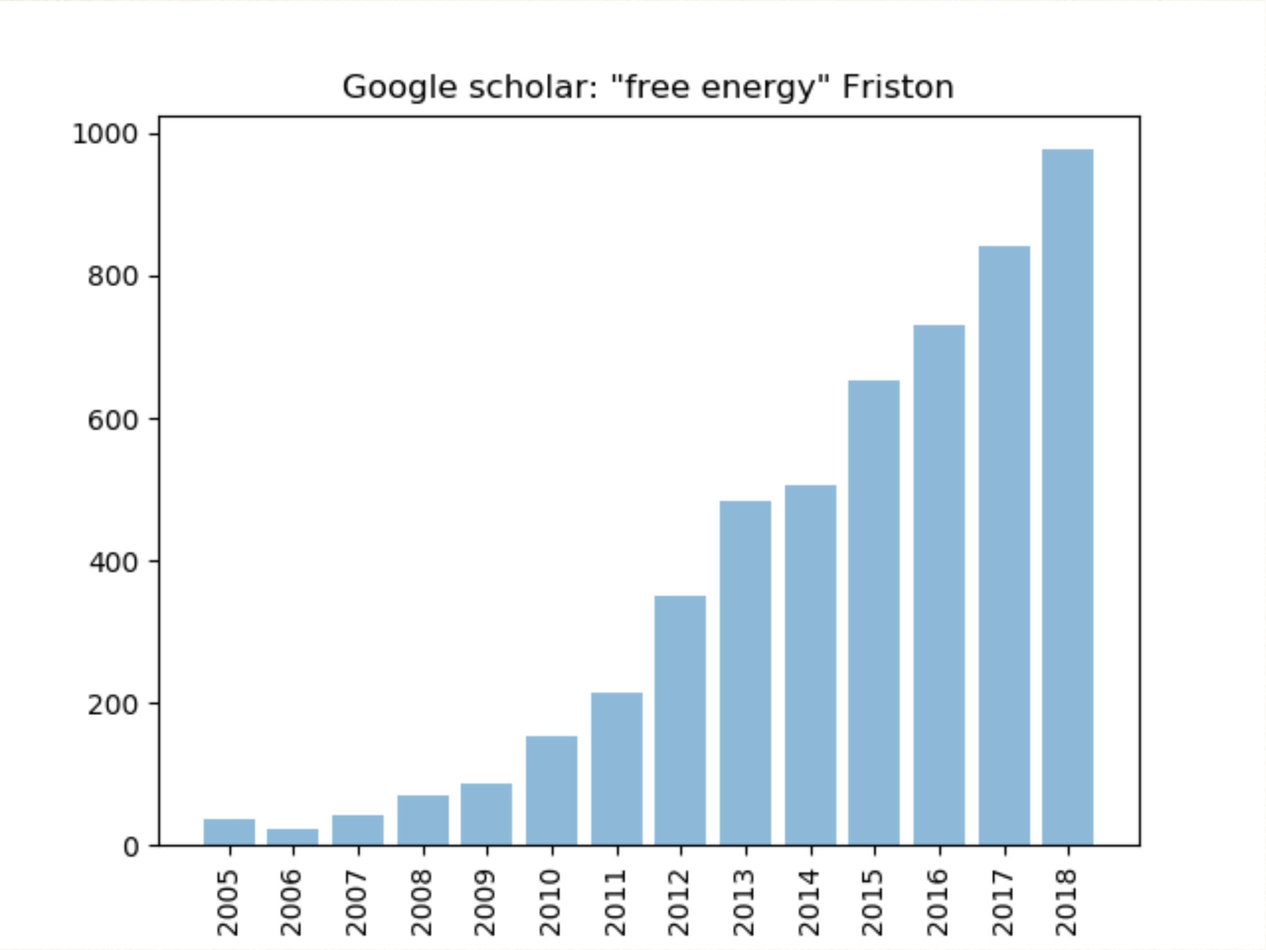
Modular minds and the separation principle

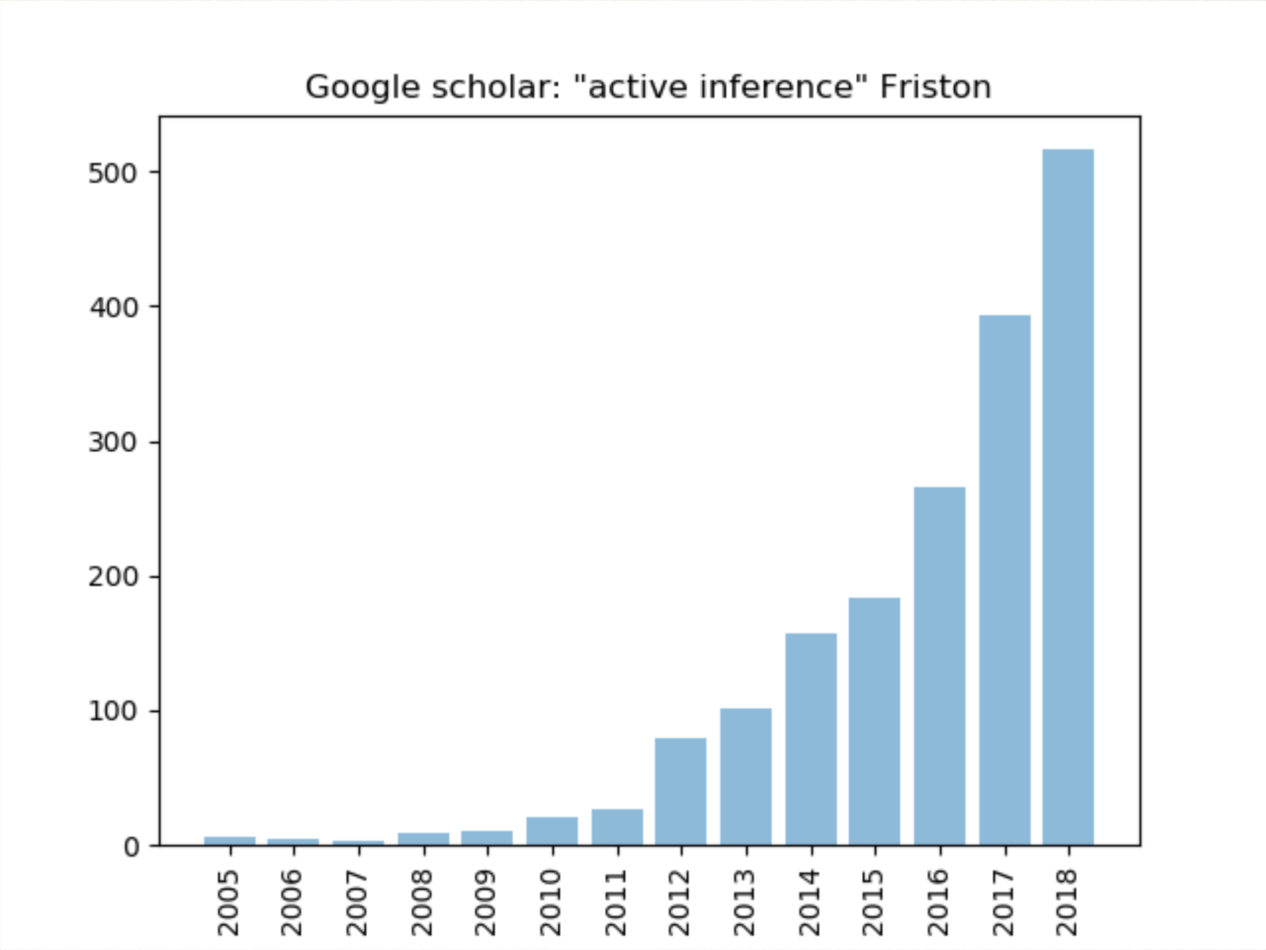
Fodor (modularity) classical sandwich	*Cog. (neuro)science*	*Control theory, separation principle*
Perception	Estimation/inference	Kalman(-Bucy) filter
↓	↓	↓
Cognition	(Complicated stuff or "just inference", à la Friston)	(Complicated stuff)
↓	↓	↓
Action	Optimal control	Linear Quadratic Regulator

The free energy principle and active inference

Karl Friston (nice profile at https://www.wired.com/story/karl-friston-free-energy-principle-artificial-intelligence/)

- Psychiatrist with a background in physical sciences
- Creator of SPM, VBM and DCM (statistical methods for the analysis of neuroimaging data)
- > 200000 citations, h-index > 200


Google scholar: "free energy" Friston
1000
800
600
400
200
0
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018


Google scholar: "active inference" Friston
500
400
300
200
100
0
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018

The free energy principle

Hypotheses: Perception and action can be described as processes of (prediction) error minimisation via Bayesian inference (**claims 1 and 2, but not 3**)

Active inference

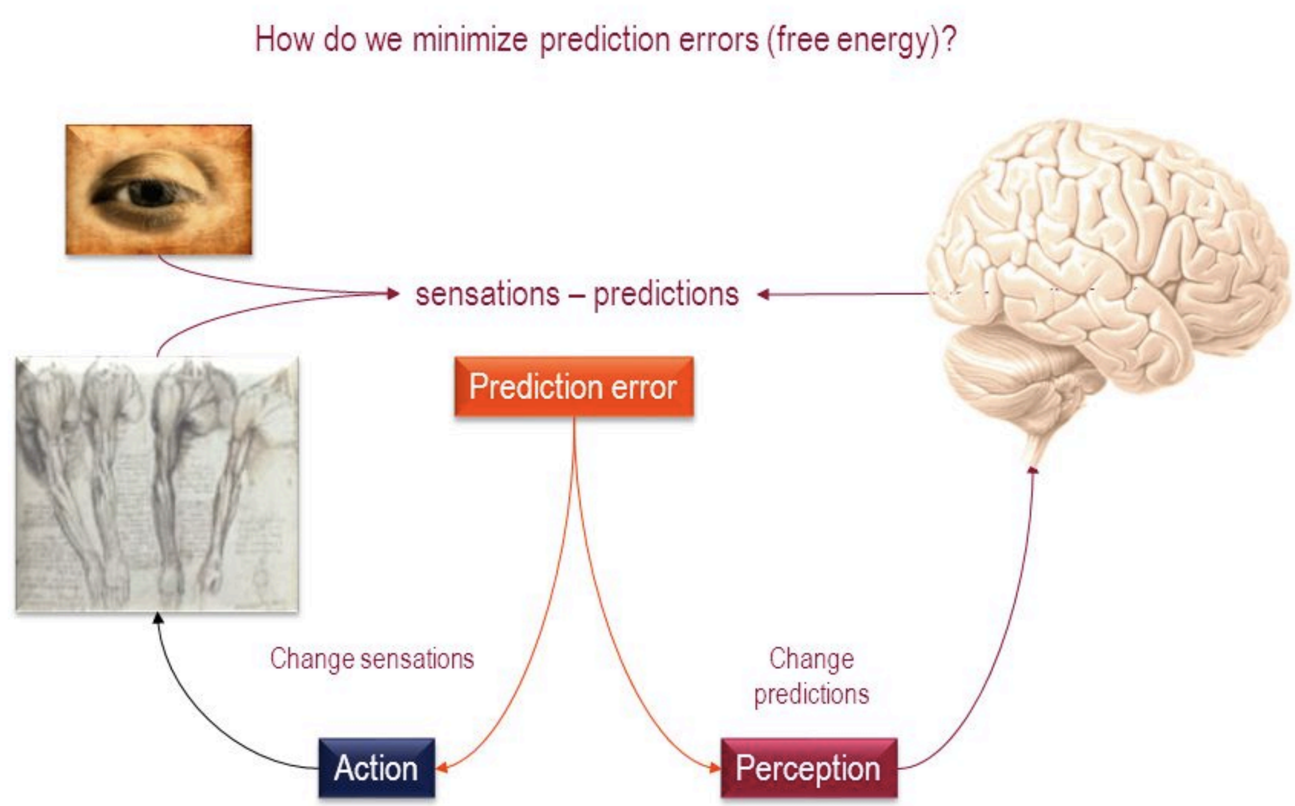


Image take from Friston, Free energy and active inference - presentation

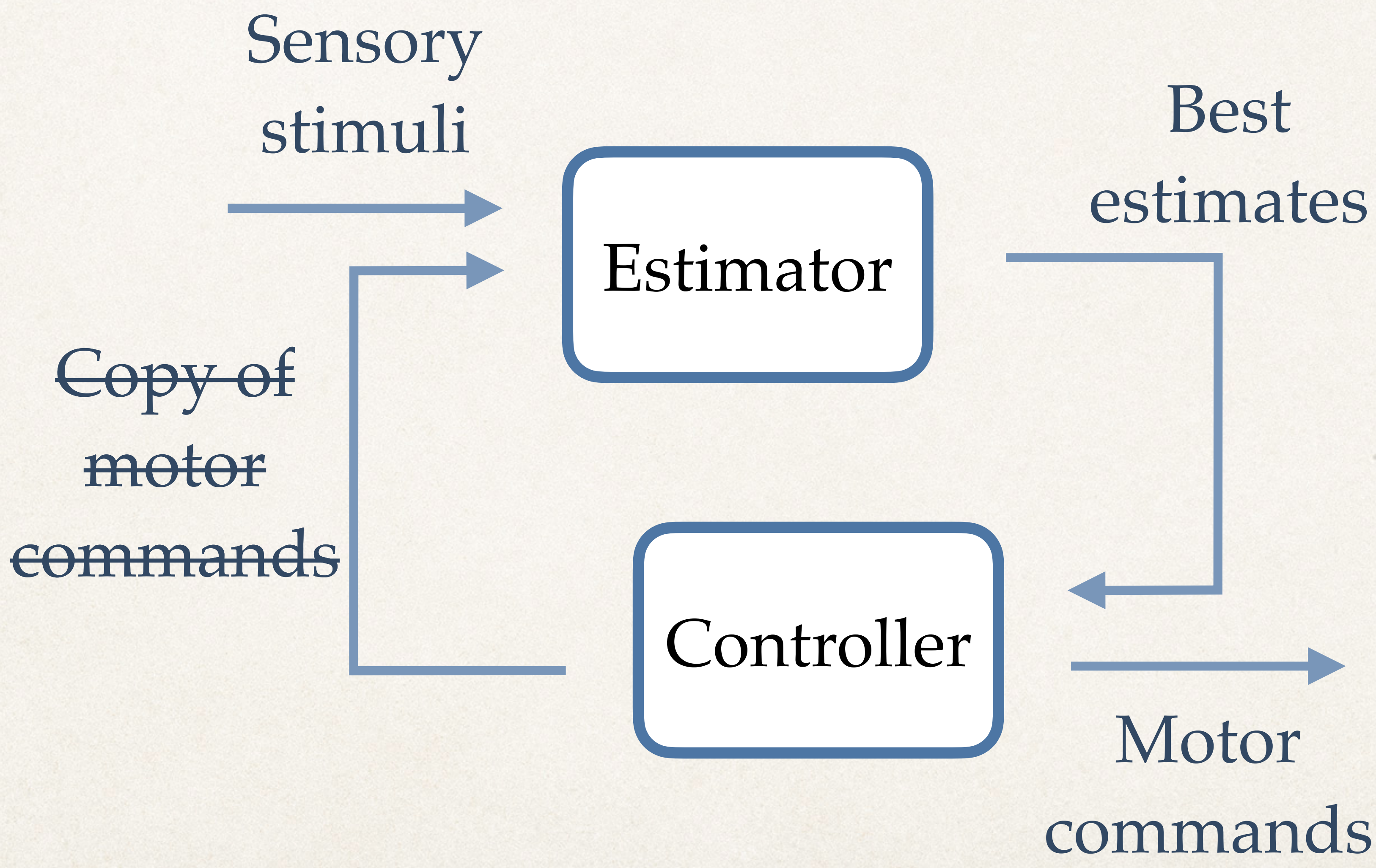
Perceptual inference

The variational free energy formulation

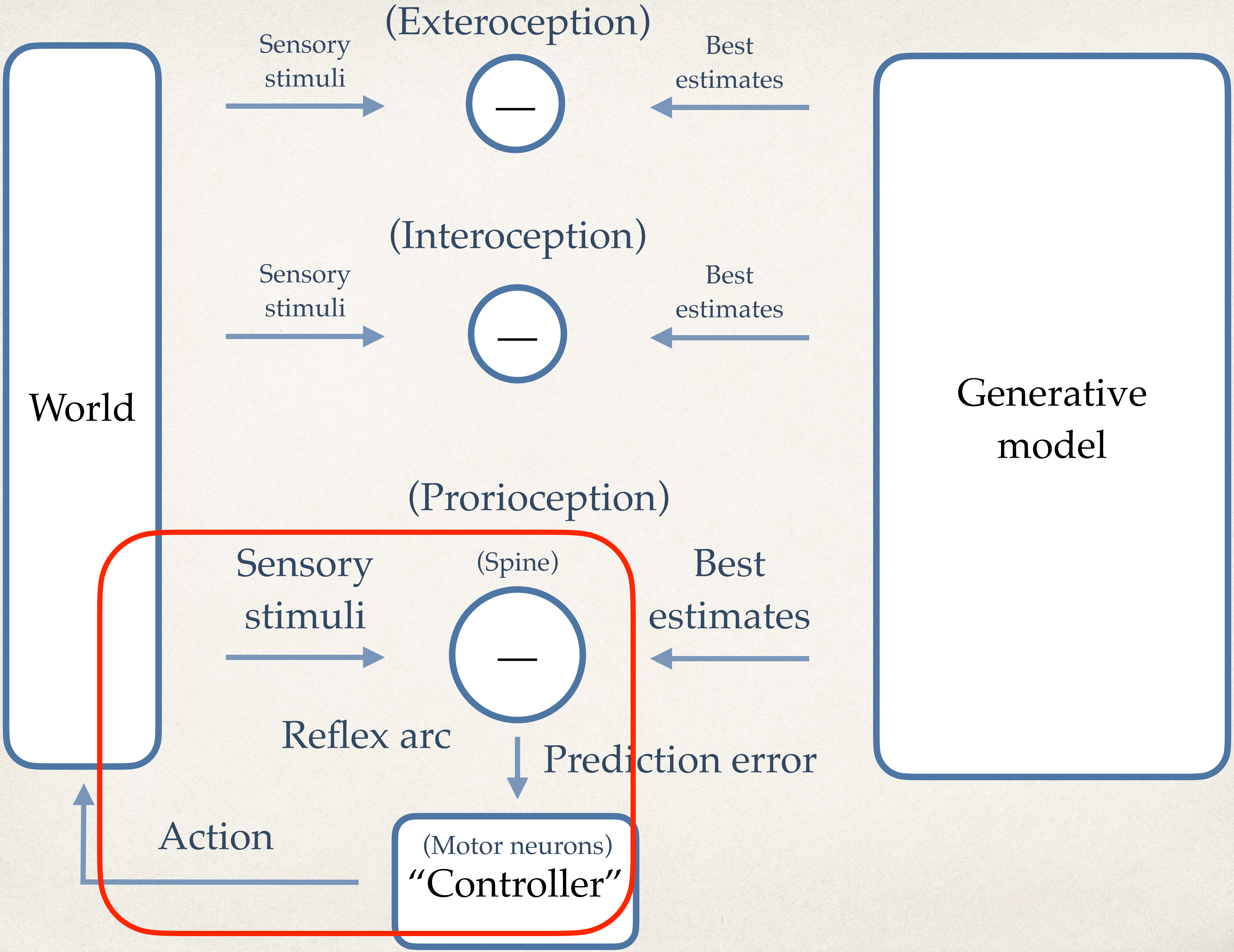
$$F \approx -\ln p(y, \vartheta)\Big|_{\vartheta=\mu_\vartheta} \qquad \vartheta = \{x, v, \theta, \gamma\}$$

- Perception $$\dot{\mu}_x = D\mu_x - \frac{\partial F}{\partial \mu_x}$$
- Action $$\dot{a} = \frac{\partial F}{\partial a} = \frac{\partial F}{\partial y}\frac{\partial y}{\partial a}$$

What about active inference?



Baltieri, Manuel, and Christopher L. Buckley. "The modularity of action and perception revisited using control theory and active inference." *Artificial Life Conference Proceedings*, 2018.

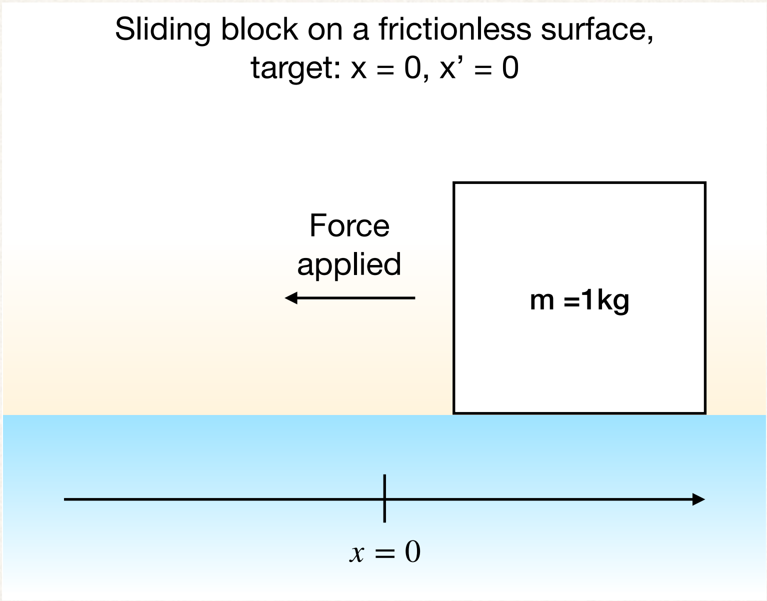

World
(Exteroception)
Sensory stimuli
—
Best estimates
(Interoception)
Sensory stimuli
—
Best estimates
Generative model
(Prorioception)
Sensory stimuli
(Spine)
—
Best estimates
Reflex arc
Prediction error
Action
(Motor neurons)
“Controller”

LQG vs active inference

Minimal model reproducing LQG vs active inference

Comparison of LQG (separation principle) and active inference equations

The role of efference copy and external forces + implications for models of behaviour



$+ z$

where matric

$A =$ $\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}$

and covarian

$\Sigma_z =$ $\begin{bmatrix} & 0\\ & \exp(-1)\end{bmatrix}$

Model of a single-joint system

LQG vs active inference

Minimal model reproducing LQG vs active inference

Comparison of LQG (separation principle) and active inference equations

The role of efference copy and external forces + implications for models of behaviour

LQG

$$
\begin{aligned}
\dot{\hat{\boldsymbol{x}}} &= A\hat{\boldsymbol{x}} + B\boldsymbol{a} + K(y - C\hat{\boldsymbol{x}}) \\
\boldsymbol{a} &= -L\hat{\boldsymbol{x}} \\
K &= PH^T(\Sigma_z)^{-1} \\
L &= R^{-1}B^TV \\
\dot{P} &= \Sigma_w + AP + PA^T - K(\Sigma_z)K^T \\
-\dot{V} &= Q + A^TV + VA - L^TRL
\end{aligned}
$$

Active inference

$$
\begin{aligned}
F \approx \frac{1}{2}\Big[&\left(\boldsymbol{y} - \hat{C}\boldsymbol{\mu_x}\right)^T \Pi_z \left(\boldsymbol{y} - \hat{C}\boldsymbol{\mu_x}\right) + \\
&+ \left(\boldsymbol{\mu}'_{\boldsymbol{x}} - \hat{A}\hat{\boldsymbol{\mu}}_{\boldsymbol{x}} - \hat{B}\boldsymbol{\mu_v}\right)^T \Pi_w \left(\boldsymbol{\mu}'_{\boldsymbol{x}} - \hat{A}\boldsymbol{\mu_x} - \hat{B}\boldsymbol{\mu_v}\right) + \\
&- \ln\left|\Pi_z\right| - \ln\left|\Pi_w\right| + (m+n)\ln 2\pi\Big]
\end{aligned}
$$

$$
\begin{aligned}
\dot{\boldsymbol{\mu}}_{\boldsymbol{x}} &= D\boldsymbol{\mu_x} - \frac{\partial F}{\partial \boldsymbol{\mu_x}} = \boldsymbol{\mu}'_{\boldsymbol{x}} + \hat{C}^T\Pi_z\left(\boldsymbol{y} - \hat{C}\boldsymbol{\mu_x}\right) + \\
&\quad + \hat{A}^T\Pi_w\left(\boldsymbol{\mu}'_{\boldsymbol{x}} - \hat{A}\boldsymbol{\mu_x} - \hat{B}\boldsymbol{\mu_v}\right) \\
\dot{\boldsymbol{\mu}}'_{\boldsymbol{x}} &= D\boldsymbol{\mu}'_{\boldsymbol{x}} - \frac{\partial F}{\partial \boldsymbol{\mu}'_{\boldsymbol{x}}} = \boldsymbol{\mu}''_{\boldsymbol{x}} - \Pi_w\left(\boldsymbol{\mu}'_{\boldsymbol{x}} - \hat{A}\boldsymbol{\mu_x} - \hat{B}\boldsymbol{\mu_v}\right)
\end{aligned}
$$

$$
\dot{\boldsymbol{a}} = -\frac{\partial F}{\partial \boldsymbol{a}} = -\frac{\partial F}{\partial \boldsymbol{y}}\frac{\partial \boldsymbol{y}}{\partial \boldsymbol{a}} = -\frac{\partial \boldsymbol{y}}{\partial \boldsymbol{a}}^T \Pi_z\left(\boldsymbol{y} - \hat{C}\boldsymbol{\mu_x}\right)
$$

The actual differences

Standard formulation (separation principle, LQG)

$$\dot{x} = Ax + Bu + w$$
$$y = Cx + z$$

u : external, or exogenous, inputs

What are the inputs?

?

$$\dot{x} = Ax + \cancel{Ba + I} + w$$
$$y = Cx + z$$

$u = a + I$

a : a copy of motor actions, efferent copy

I : (actual) *external* inputs

$$\dot{x} = Ax + Bv + w$$
$$y = Cx + z$$

v : priors, *desired trajectories* for movement

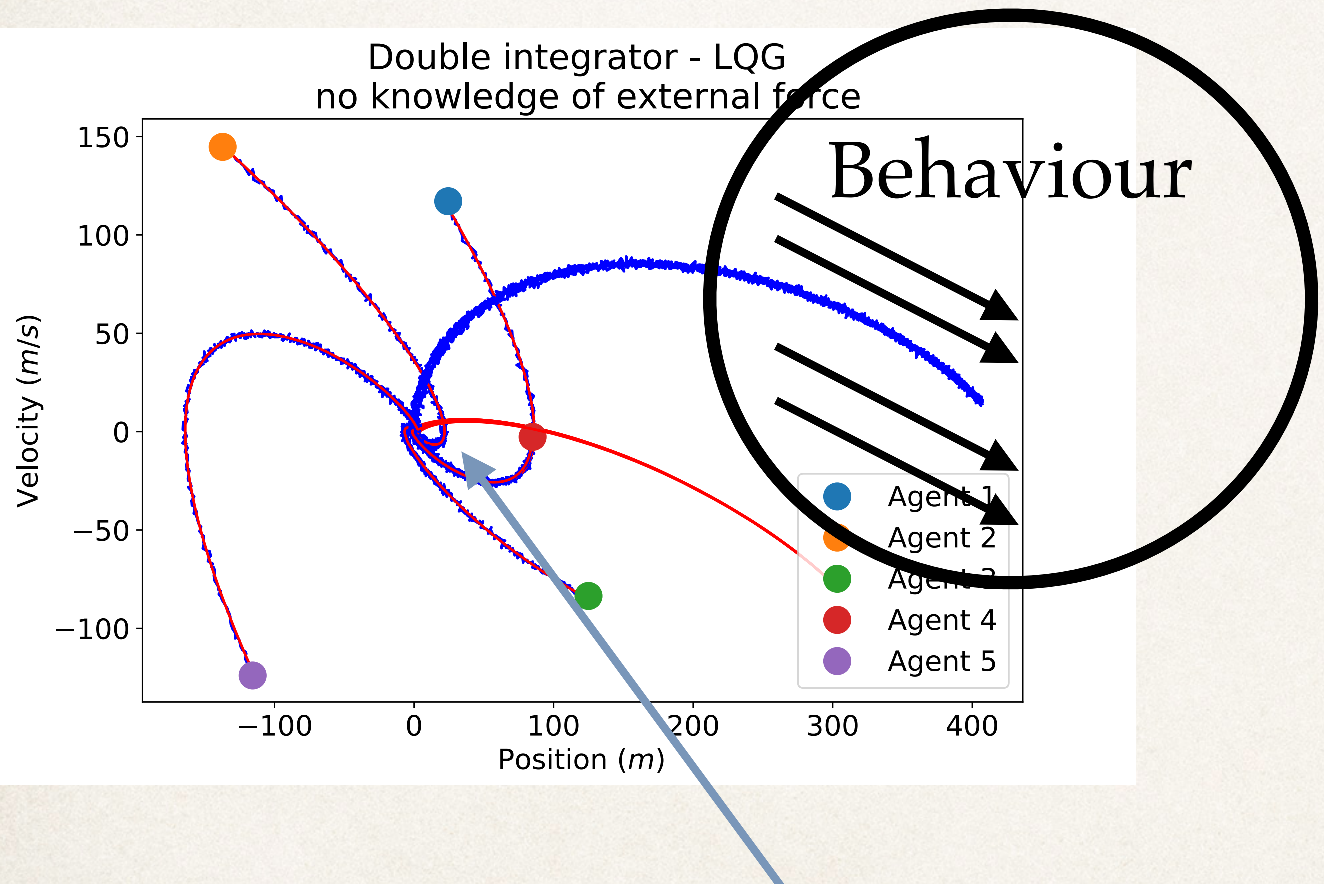
LQG vs active inference

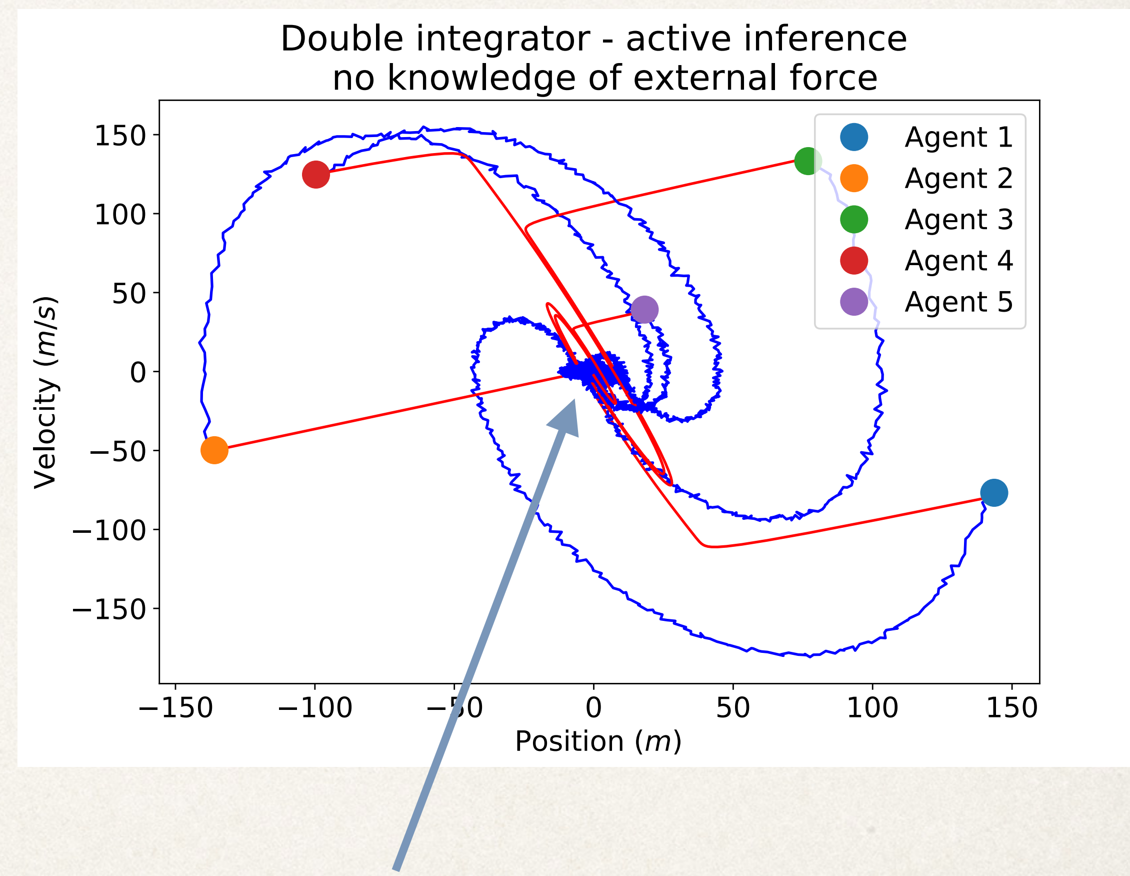
Minimal model reproducing LQG vs active inference

Comparison of LQG (separation principle) and active inference equations

The role of efference copy and external forces + implications for models of behaviour

LQG vs active inference, with external disturbance (e.g., wind)





Target

Target

Summary

- We like modularity (biology, engineering, software production, economics, etc.), and modularity is good
- Modularity in cognitive science and neuroscience is a standard assumption (classical sandwich, input-output architectures), but not very realistic for models of cognition and behaviour
- Modularity aligns well (?) with the separation principle of control theory
- The separation principle is only but a linear approximation of feedback systems and copies of motor signals may not be biological plausible
- Active inference drops the separation assumption and applies Bayesian inference and optimal control to more general problems (e.g., unknown forces that affect observations) with no need for classical modularity

Acknowledgements

- Christopher L. Buckley

- Olaf Witkowsky / Nathaniel Virgo

- Simon McGregor / Keisuke Suzuki / Anil Seth / Warrick Roseboom / Lionel Barnett (University of Sussex)
- Taro Toyoizumi (Riken CBS)
- Hideaki Shimazaki (University of Kyoto)
- Takashi Ikegami (University of Tokyo)
- Martin Biehl (Araya Inc.)
- Nicholas Guttenberg (ELSI)

Contacts:

Email: manuel.baltieri@gmail.com

Twitter: @manuelbaltieri