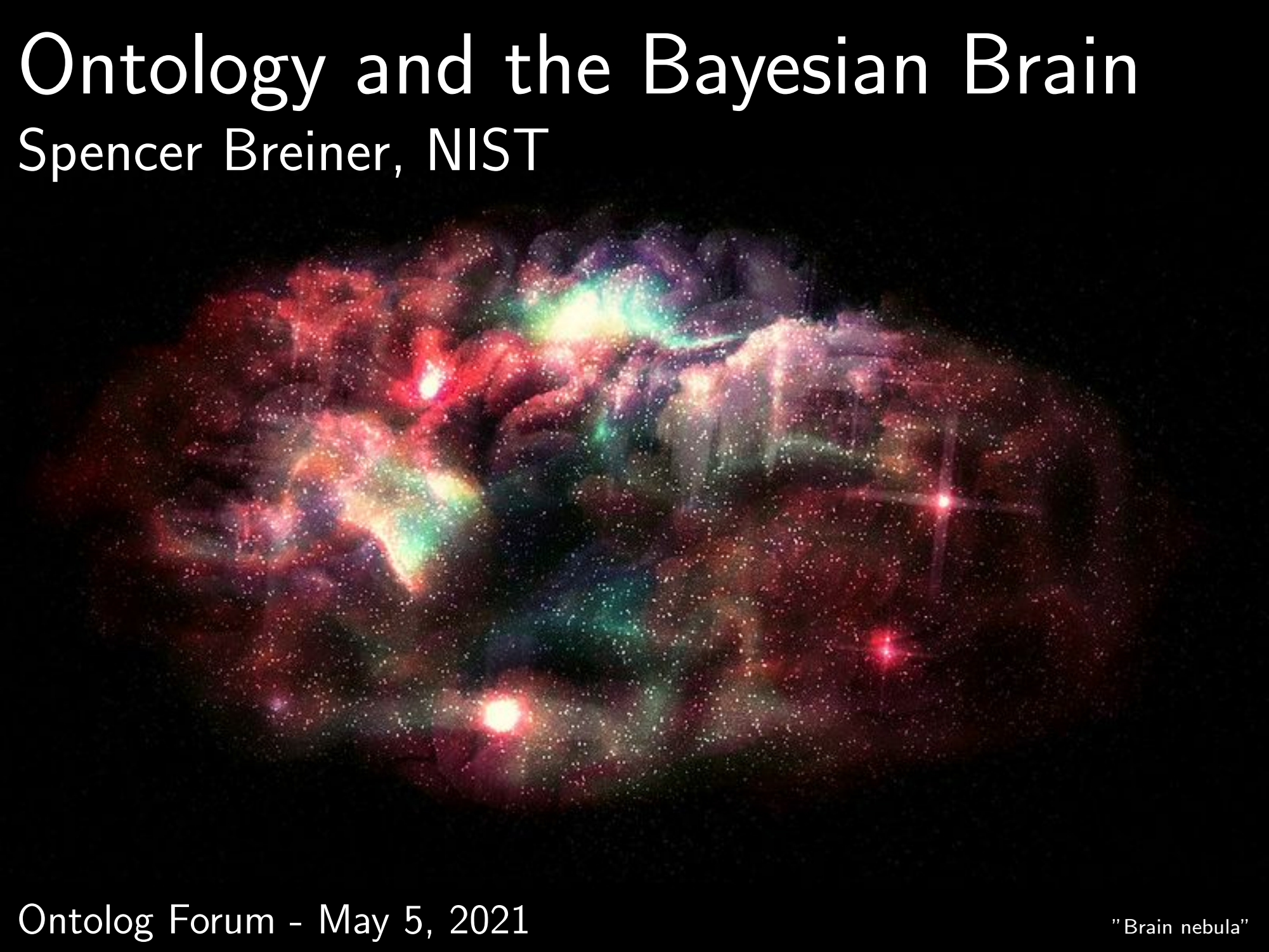
Ontology and the Bayesian Brain

Spencer Breiner, NIST

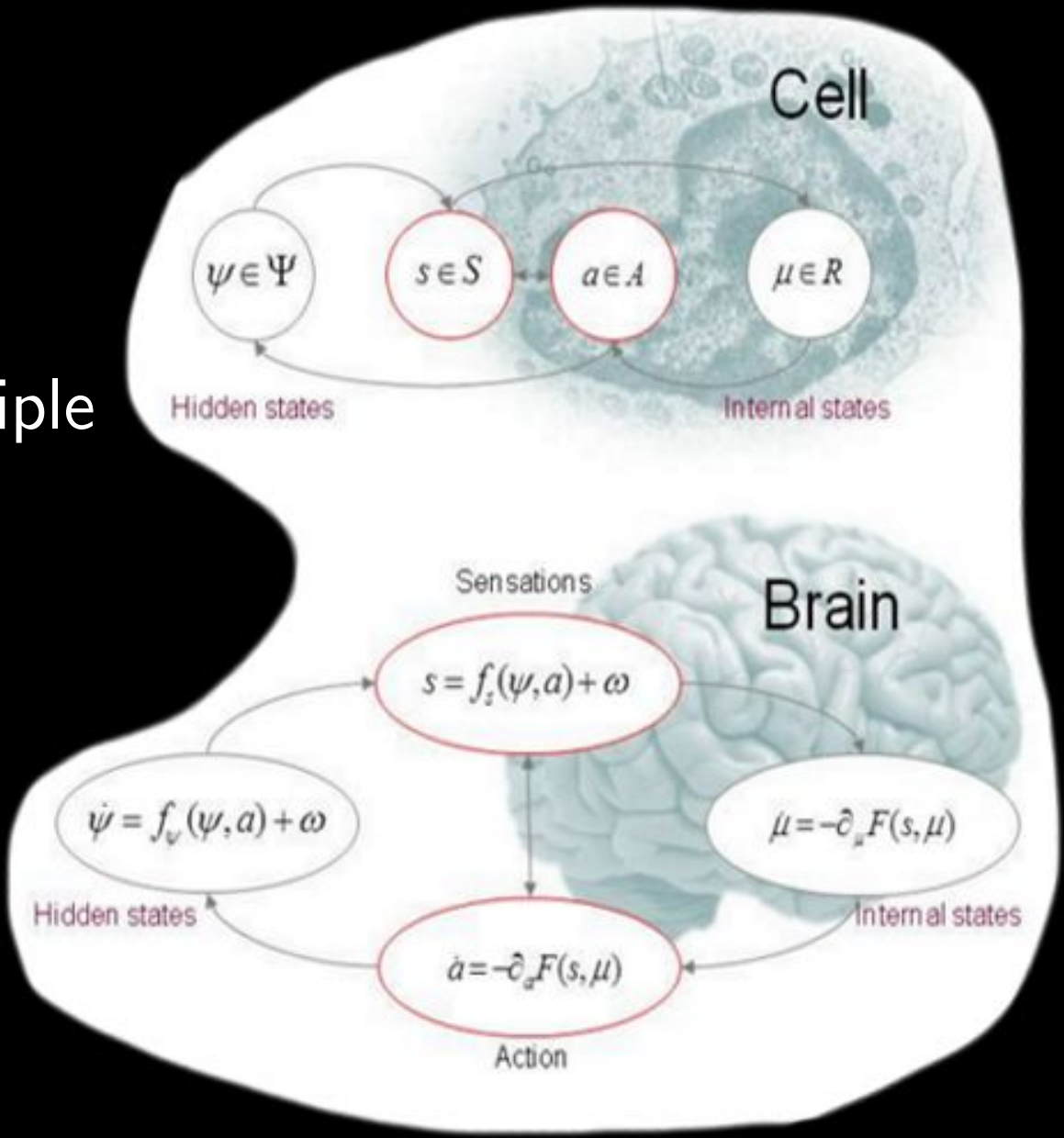


Ontolog Forum - May 5, 2021

"Brain nebula"

The plan for today:

- The Bayesian brain
- The free energy principle
- Sensing & acting
- Bayesian ontologies?

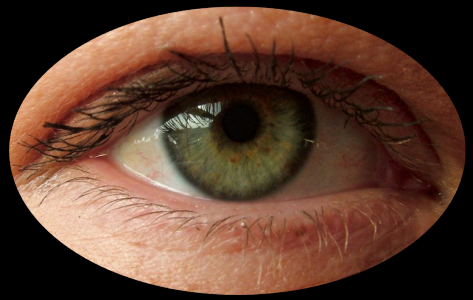


Some caveats:

- More neuro-, less symbolic (no NLP)
- I'm still learning this stuff
- Theory $\gg$ evidence/implementation

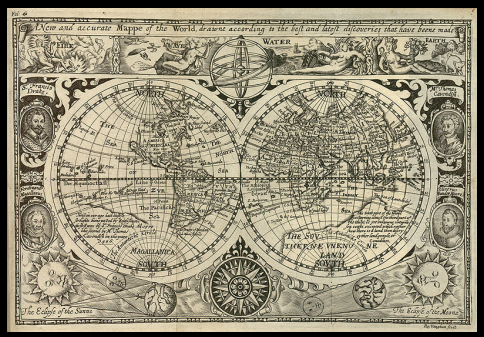
The Bayesian Brain

The sense-action loop









Unconscious Inference



"What we see is the solution to a computational problem, our brains compute the most likely causes from the photon absorptions within our eyes."

- Helmholtz (?)

"...[T]he observed action on our senses enables us to form an idea as to the possible cause of this action...."

- Helmholtz (1867)

Predictive Coding

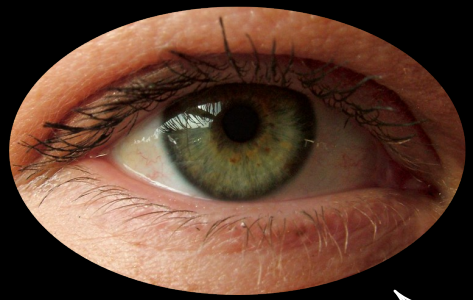




"[F]eedback connections from a higher- to a lower-order visual cortical area carry predictions of lower-level neural activities, whereas the feedforward connections carry the residual errors between the predictions and the actual lower-level activities."

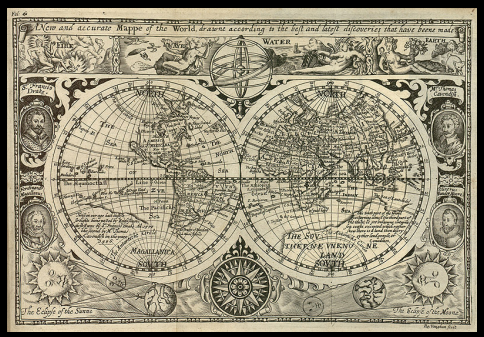
- Rao & Ballard (1999)

The sense-action loop (c. 1999)









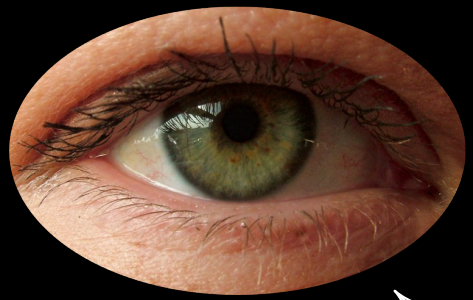
Active Inference

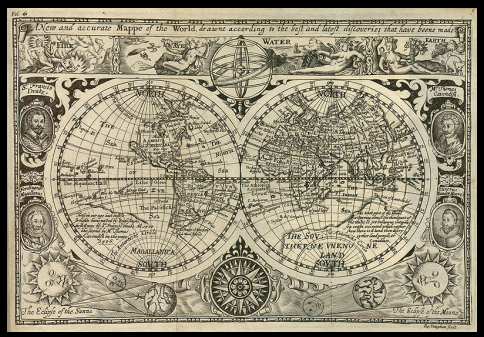


"[I]n simulating their little piece of synthetic cortex, neuronal dynamics and connectivity optimized the same energy or cost function. I remember reading the methods section several times to convince myself that they could explain all of this functional anatomy and detailed neurophysiology with just one energy function."

- Friston (2018)
on Ballard & Rao (1999)

The sense-action loop (c. 2005)

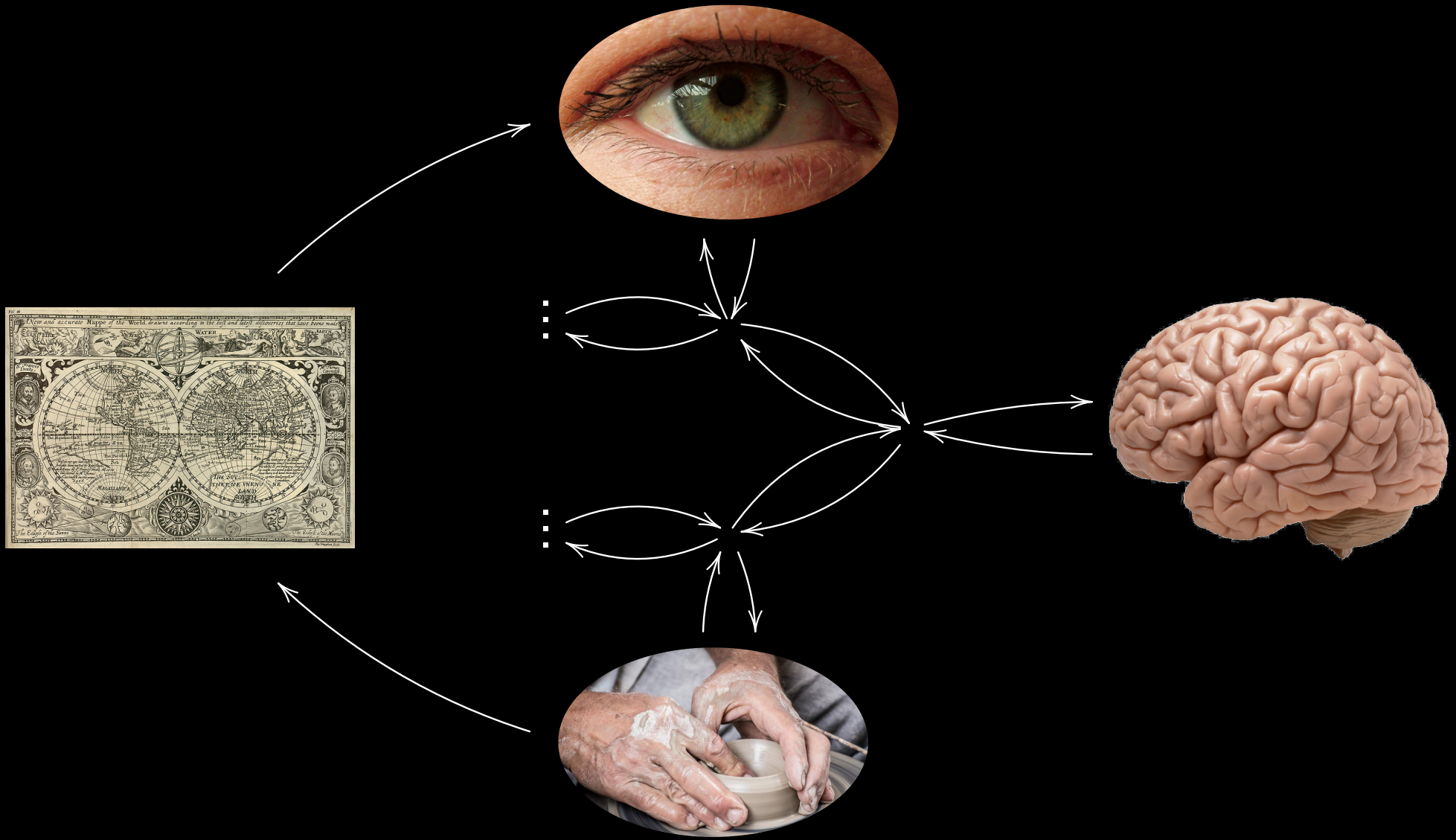








The sense-action loop (c. 2020)



The Free-Energy Principle

In words

Self-organizing systems infer and act by minimizing the variational free energy of two probabilistic distributions: a prior on the environment and a generative model of sensory input.

- Homeostasis is a region of sensory space.
- Free energy places a (tractable) upper bound on distance from this region.
- For perception, FEM entails coherence of expectations and experience.
- For action, FEM entails goal-directed behavior and information-seeking.

Main characters

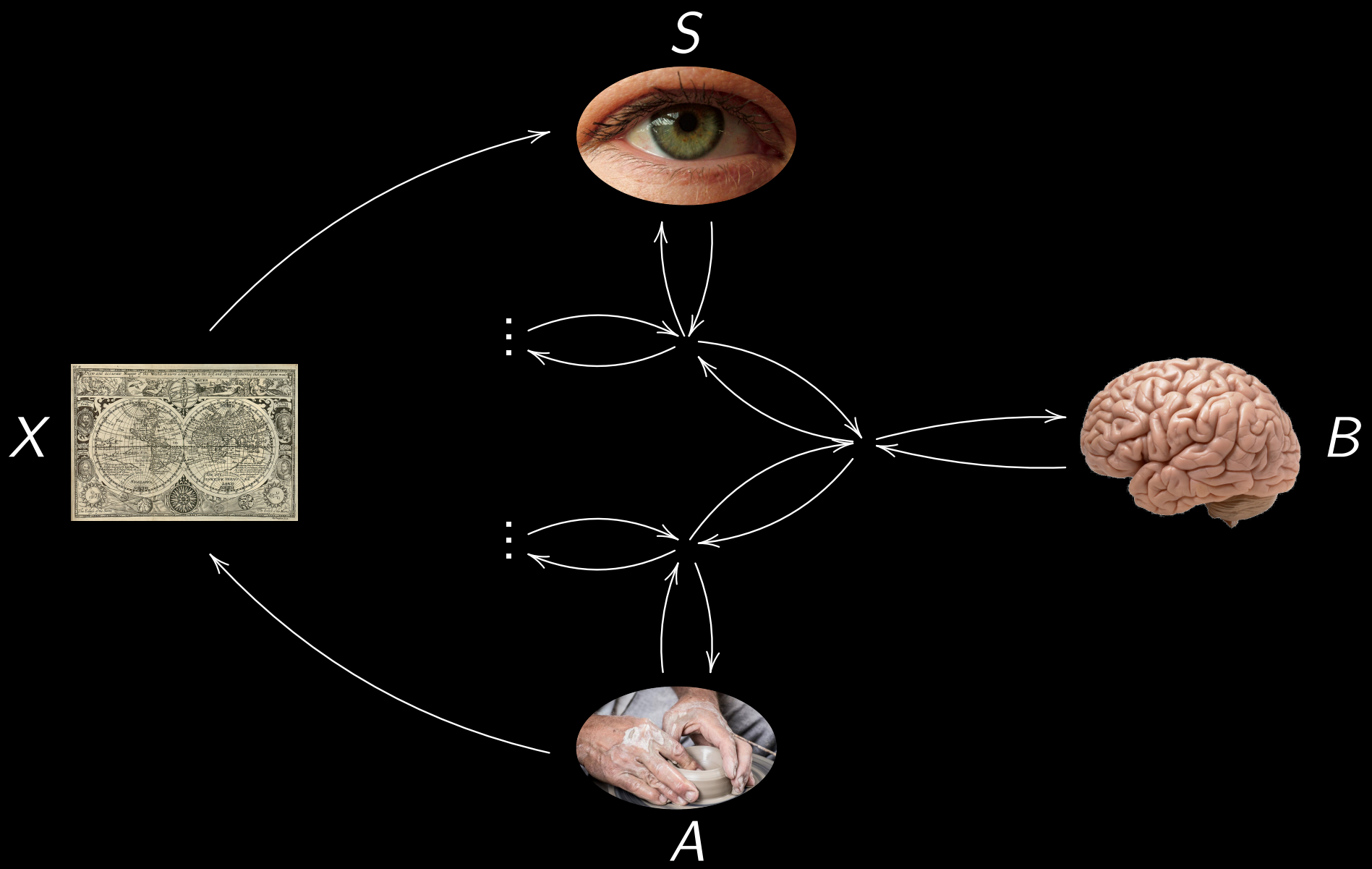
Env. states	Sensory states	Action states	Brain states
$x \in X$	$s \in S$	$a \in A$	$b \in B$

Recognition density	Generative density
$R(x\|b)$	$G(x, s\|a)$

Dynamics	Inference
$X \times A \xrightarrow{\texttt{dyn}} X \times S$	$B \times S \xrightarrow{\texttt{infr}} B \times A$
$\texttt{evol}, \texttt{obs}$	$\texttt{update}, \texttt{act}$

Claim: $\texttt{infr}(b, s) = \arg\min_{b,a} F(b, s, a)$

Main Characters



Main characters

Env. states	Sensory states	Action states	Brain states
$x \in X$	$s \in S$	$a \in A$	$b \in B$

Recognition density

$$R(x|b)$$

Generative density

$$G(x, s|a)$$

Dynamics

$$X \times A \xrightarrow{\texttt{dyn}} X \times S$$

$\texttt{evol}$,$\texttt{obs}$

Inference

$$B \times S \xrightarrow{\texttt{infr}} B \times A$$

$\texttt{update}$, $\texttt{act}$

Claim: $\texttt{infr}(b, s) = \arg\min_{b,a} F(b, s, a)$

Surprise!

Given a distribution $p = \langle p_x \rangle$:

$$!p_x = \texttt{surprise}(p, x) := -\log p_x$$



$$[0, 1] \overset{!}{\cong} [\infty, 0]$$

$$p = p_1 \cdot p_2 \implies !p = !p_1 + !p_2$$

For the generative model of sensation, $!p_x \gg 1$ is bad!

Kullback-Liebler Divergence

Given two probability distributions p and q

$$D(p,q) := E_p\left[!p_x - !q_x\right] = \sum_x p_x \log \frac{p_x}{q_x}$$

Key property:

$$D(p,q) \geq 0 \;\&\; D(p,q) = 0 \Longrightarrow p = q \text{ a.e.}$$

... but *not* a metric.

$D(p,q)$ measures information gain $q \to p$:

$$D(H, \tfrac{1}{2}/\tfrac{1}{2}) = 1 \cdot \log_2 \frac{1}{1/2} + 0 \cdot \log_2 \frac{0}{1/2} = 1 \text{ bit}$$

Free Energy

$$F(b,s,a) = \sum_x R(x|b) \log \frac{R(x|b)}{G(x,s|a)}$$

$G(x,s|a)=G(x|s,a)\cdot G(s|a)$

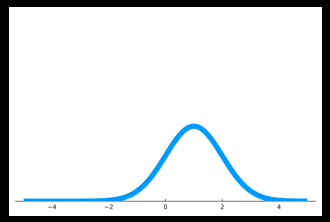
$$= \overbrace{!G(s|a)}^{\text{Surprise}} + \overbrace{D\big(R(x|b), G(x|s,a)\big)}^{\text{Conditional divergence}}$$

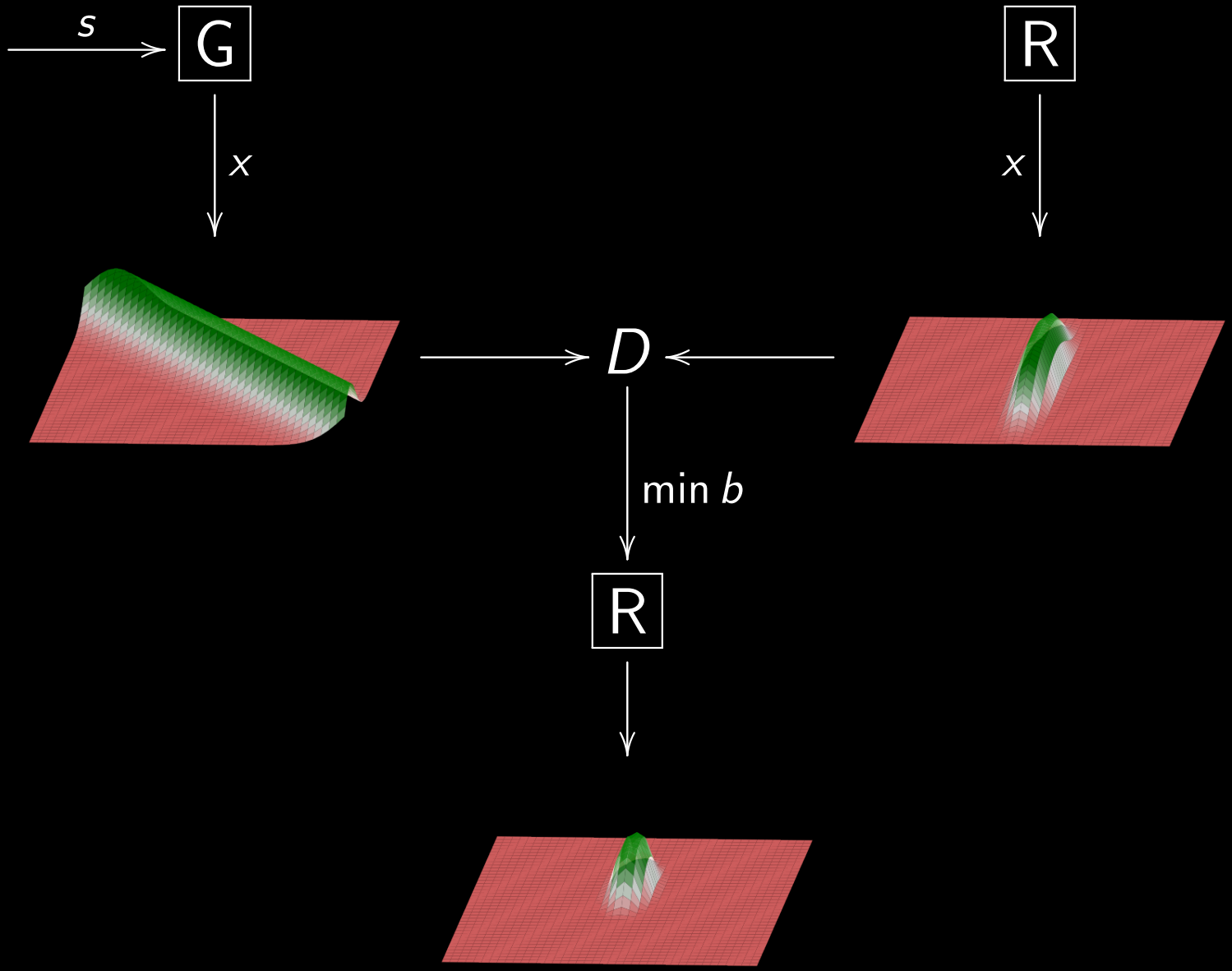
$G(x,s|a)=G(s|x,a)\cdot G(x|a)$

$$= \overbrace{E_R\big[!G(s|x,a)\big]}^{\text{Inaccuracy}} + \overbrace{D\big(R(x|b), G(x|a)\big)}^{\text{Complexity}}$$

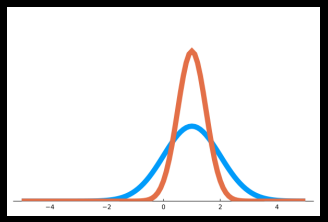
Sensing & acting

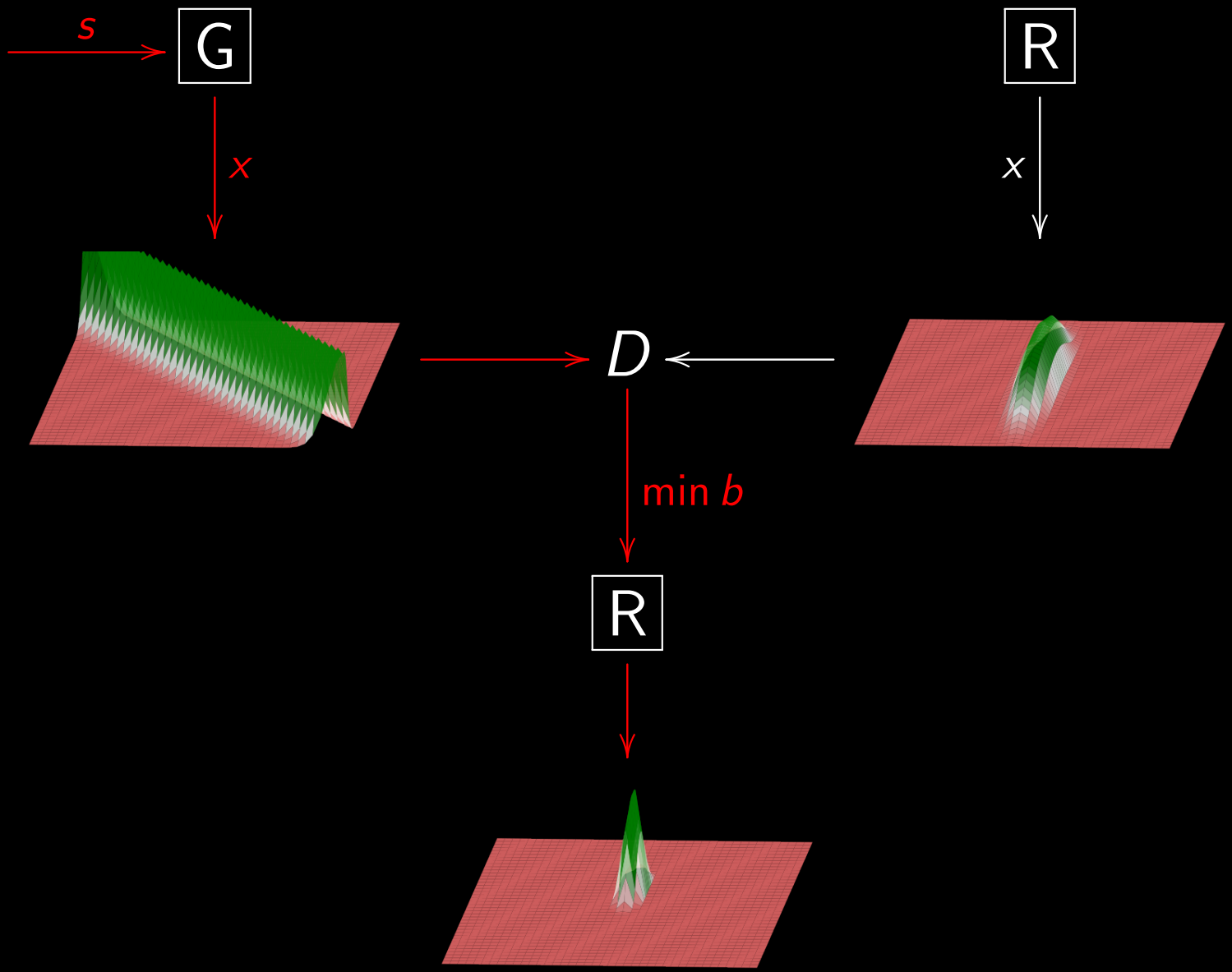
Predictive coding





Precision weighting 1





Precision weighting 1

$$F(b, s, a) = \sum_x R(x|b) \log \frac{R(x|b)}{G(x,s|a)}$$

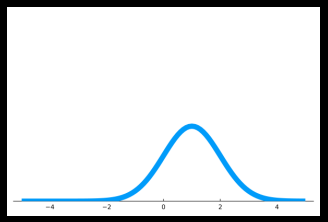
$G(x,s|a)=G(x|s,a)\cdot G(s|a)$

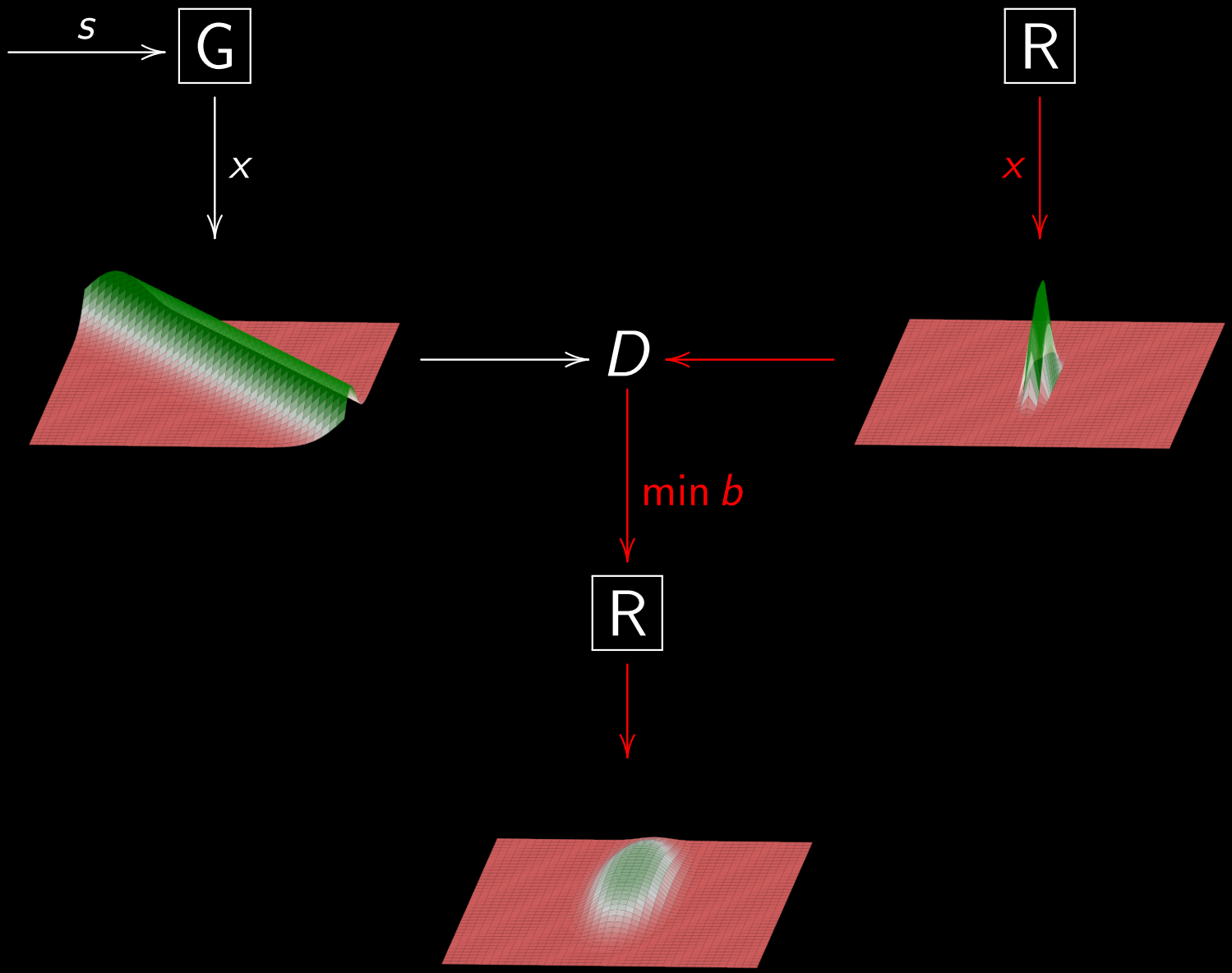
$$= \overbrace{!G(s|a)}^{\text{Surprise}} + \overbrace{D\big(R(x|b), G(x|s,a)\big)}^{\text{Conditional divergence}}$$

$G(x,s|a)=G(s|x,a)\cdot G(x|a)$

$$= \overbrace{E_R\big[!G(s|x,a)\big]}^{\text{Inaccuracy}} + \overbrace{D\big(R(x|b), G(x|a)\big)}^{\text{Complexity}}$$

Precision weighting 2





Precision weighting 2

$$F(b,s,a) = \sum_x R(x|b) \log \frac{R(x|b)}{G(x,s|a)}$$

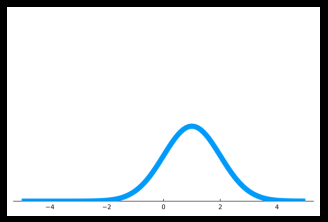
$G(x,s|a)=G(x|s,a)\cdot G(s|a)$

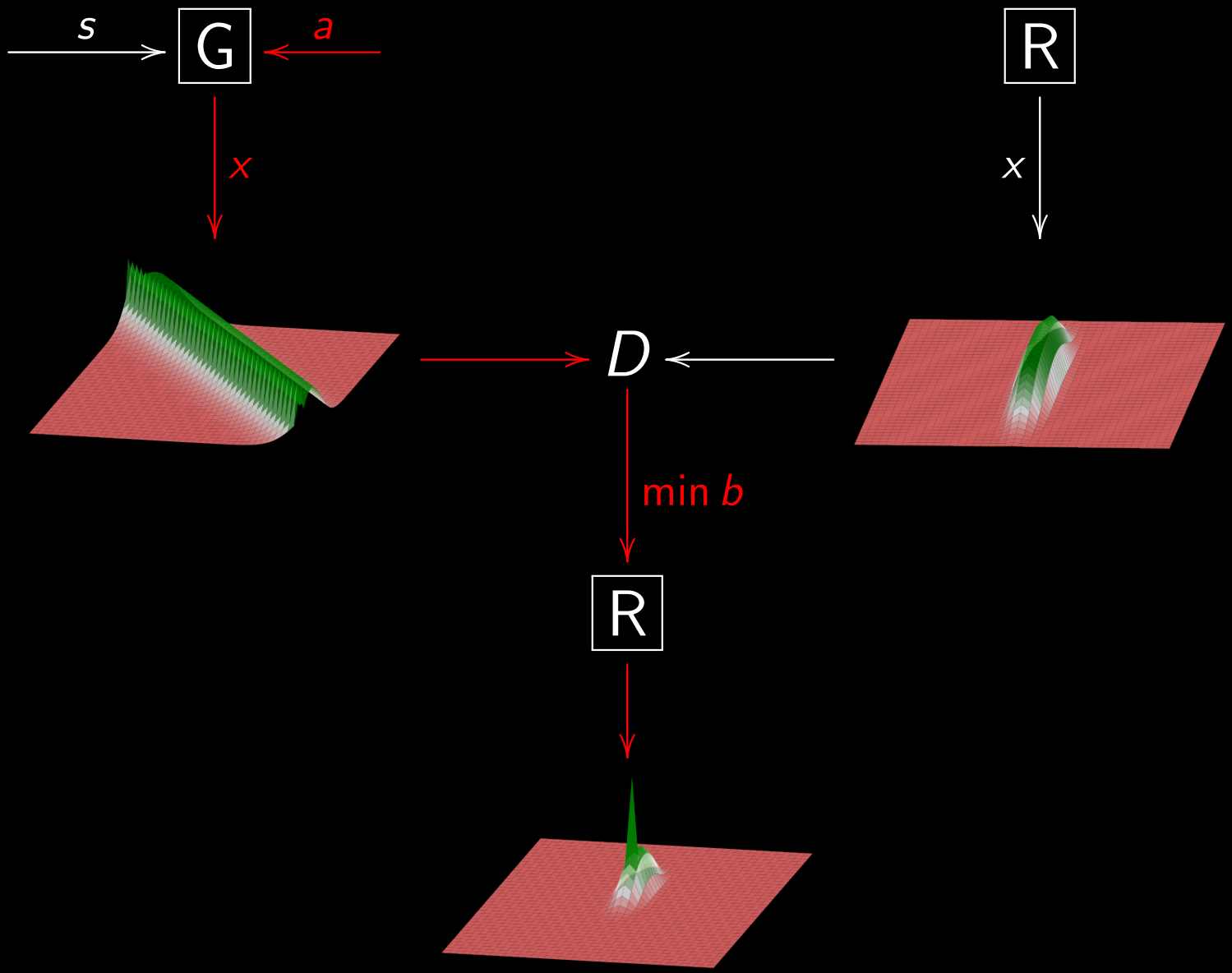
$$= \overbrace{!G(s|a)}^{\text{Surprise}} + \overbrace{D\big(R(x|b), G(x|s,a)\big)}^{\text{Conditional divergence}}$$

$G(x,s|a)=G(s|x,a)\cdot G(x|a)$

$$= \overbrace{E_R\big[!G(s|x,a)\big]}^{\text{Inaccuracy}} + \overbrace{D\big(R(x|b), G(x|a)\big)}^{\text{Complexity}}$$

Active inference





Active inference

$$F(b, s, a) = \sum_x R(x|b) \log \frac{R(x|b)}{G(x,s|a)}$$

$$\overset{G(x,s|a)=G(x|s,a)\cdot G(s|a)}{=} \overbrace{!G(s|a)}^{\text{Surprise}} + \overbrace{D\big(R(x|b), G(x|s,a)\big)}^{\text{Conditional divergence}}$$

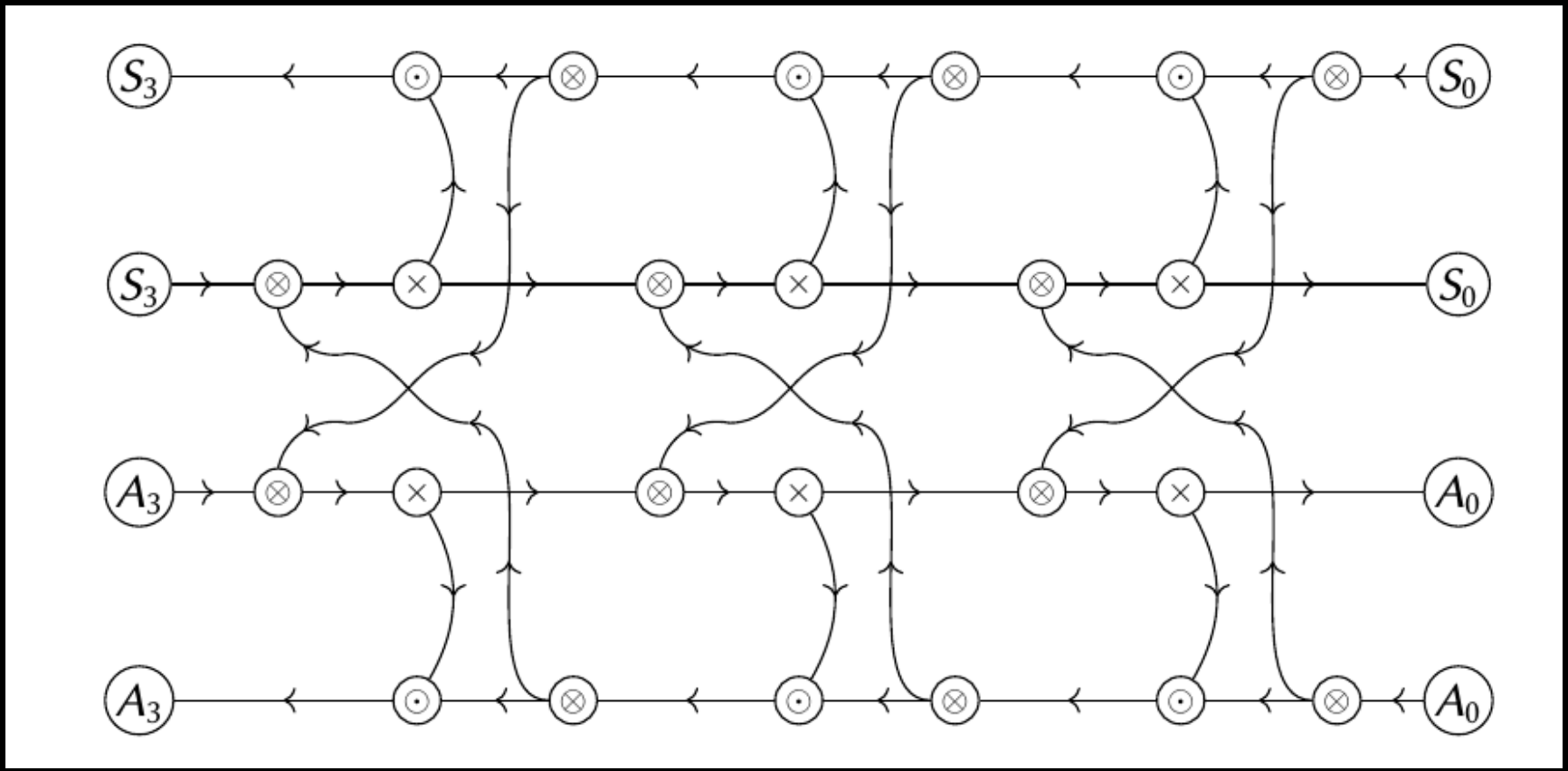
$$\overset{G(x,s|a)=G(s|x,a)\cdot G(x|a)}{=} \overbrace{E_R\big[!G(s|x,a)\big]}^{\text{Inaccuracy}} + \overbrace{D\big(R(x|b), G(x|a)\big)}^{\text{Complexity}}$$

(So many) More details

- Everything is dynamical.
- Non-smooth distributions.
- Hierarchical structure.
- Policies & future actions.
- What's a model?
- Minimization over what?

The obligatory plug

Toby St Clere Smithe (Oxford/Topos) has been working to sort out the mathematics.



Bayesian Ontologies (?)

The takeaway

Human understanding is probabilistic.

To reflect our best understanding of the world, ontologies must be, too.

But how?

- Use ontologies to construct probabilities (BayesOWL).
- Use probability to construct ontologies (AFOG).
- Probability on top of ontology (PR-OWL).
- Ontology on top of probability (BEL).
- Probability to link ontologies (GLUE)

A few observations

- Certainty is the special case.
- Probabilities are dynamic.
- Improvement $\gg$ Quality (long-term).
- We learn through conflict.

My two cents:

Need a meta-theory for probabilistic ontology that is

compositional and **probability-native**.

Thanks!

(& refs)

References

Papers

- Carvalho, Laskey & Costa, PR-OWL:a language for defining probabilistic ontologies
- Ceylan & Peñalosa, The Bayesian Ontology Language $\mathcal{BEL}$
- Friston & Stephan, Free-energy and the brain
- Friston, Does predictive coding have a future?
- Friston, Daunizeau & Kiebel, Reinforcement Learning or Active Inference?
- Helmholtz (Southall, trans.), Treatise on Physiological Optics, v. 3
- Pan, Ding, Yu & Ping, A Bayesian Network Approach to Ontology Mapping
- Rao & Ballard, Predictive Coding in the visual cortex
- Spratling, A Review of Predictive Coding Algorithms
- Smithe, Cyber kittens, or Some First Steps Towards Categorical Cybernetics
- Smithe, The Compositional Structure of Active Inference
- Solopchuk, Tutorial on Active Inference
- Wang, et al., Formal Ontology Generation by deep machine learning

References

Images

- "Brain nebula", ezhikoff, CC-BY-2.0
- "Markov Blanket", Kfriston, CC-BY-SA-3.0
- "Suzanne's eye", Eelco, CC-BY-NC
- "human brain on white background", _DJ_, CC-BY-SA
- "Clay Pottery Hands", LubosHouska, Pixabay
- "Hermann von Helmholtz", Wikimedia, public domain
- "Rajesh Rao", Joe Mabel, CC-SA-4.0
- "Dana Ballard", © Dana Ballard
- "Karl Friston", © Kate Peters, Wired
- "bad surprise", Brandy Bora, CC-BY
- "Mappe of the World", Krause Collection, Library of Congress
- "Hierarchical Active Inference Game", Toby Smithe