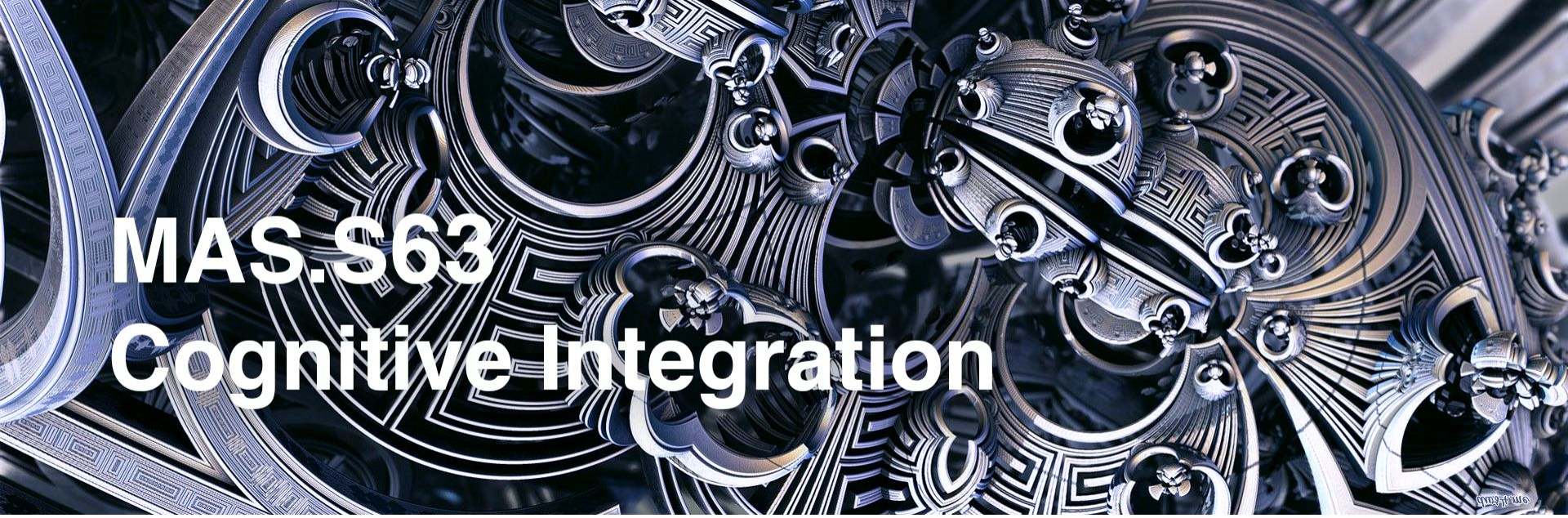


MAS.S63
Cognitive Integration

Spring Term 2016
New Destinations in Artificial Intelligence

joscha@mit.edu adam.h.
marblestone@gmail.com

Questions

What are minds?

Which is the best discipline to study the mind?

What is the right methodology to study the mind?

Which parts of minds are inaccessible to science?

Which parts of minds are already explained well?

Which parts should we focus on?

When can we hope to have a complete theory of mind and intelligence?

Overview

- Minds as information processing systems
- The basic architecture of the human mind
- The design space of cognitive architectures
- Mental representation
- From perception to thought
- Mechanisms of attention
- Imagination and creativity
- Social cognition
- Learning

Organization

- Seminar, 12 sessions
- weekly meetings, discussion & presentations
- practical implementations/projects, starting Friday

Grading: project & presentation required

- get in touch ASAP (joscha@mit.edu)
- weight 1:2:2 (participation, presentation, paper)

- structure & material is open to suggestions

Sessions: see futureai.media.mit.edu

2/8 Introduction

2/16 (TUESDAY!): Symbolic vs. connectionist AI

2/22: Theories of perception, representation, symbol grounding

2/29: Learning

3/7: MicroPsi

3/14: Social cognition, theory of mind

3/28: Cortical organization

4/4: Computational models of cortical function

4/11: Imagination and creativity

4/25: Spaun

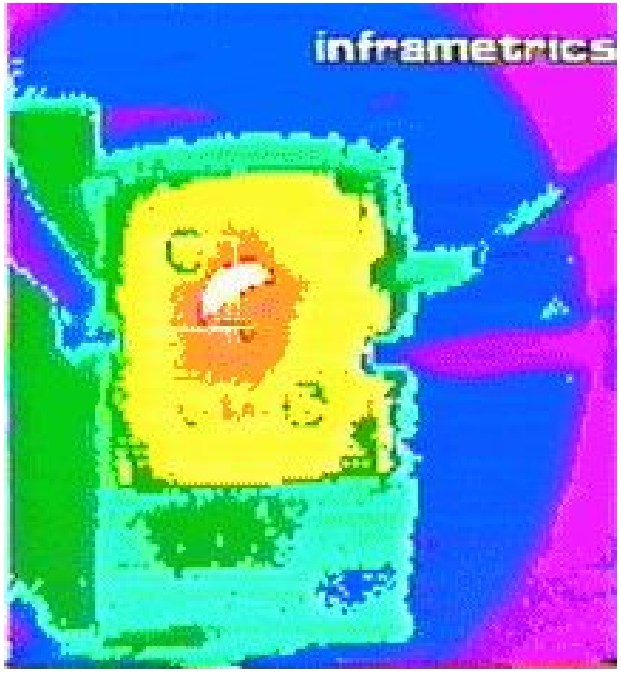
5/2: Leabra

5/9: Closing Discussion. The design space of cognitive architectures

Mind as Machine

Functionalist modeling

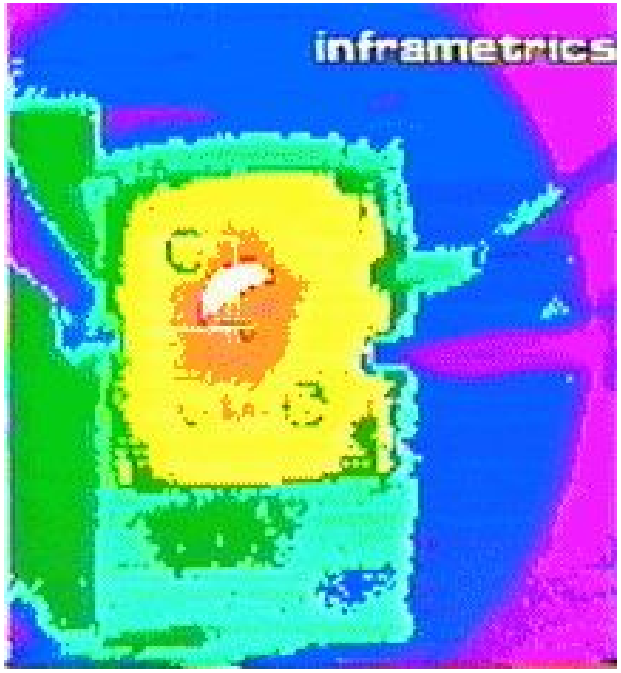
1. Build whole, **functionalist** architectures



(infrared) imaging of combustion engine

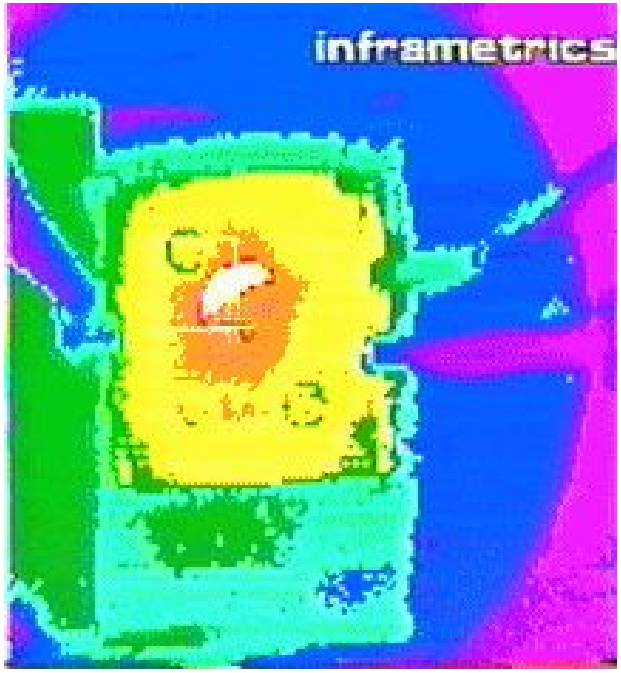
Functionalist modeling

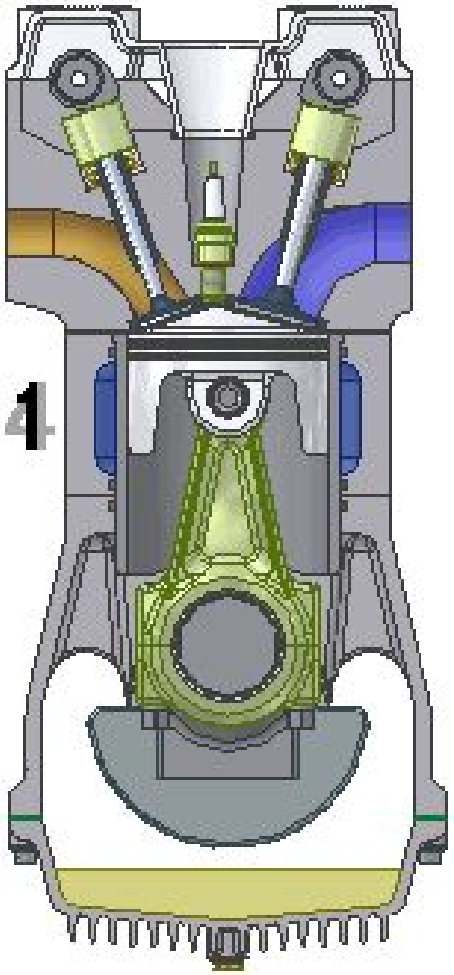
1. Build whole, **functionalist** architectures



(infrared) imaging of combustion engine

1. Build whole, **functionalist** architectures



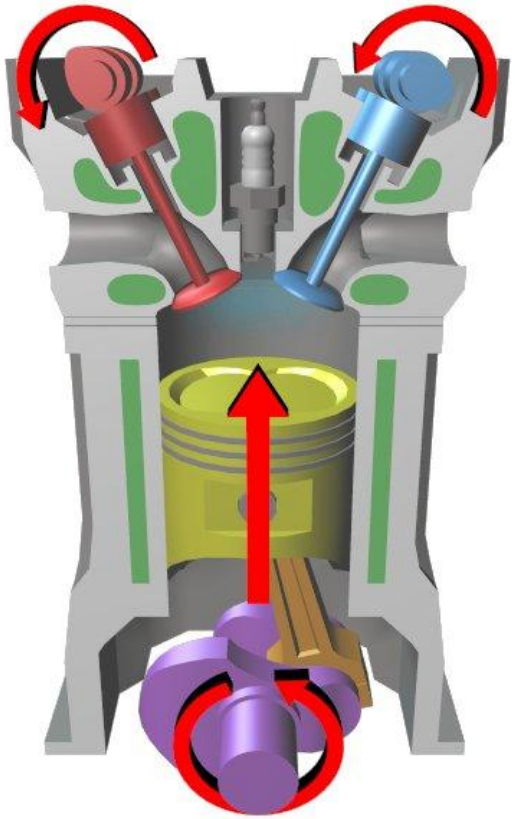


Functionalist modeling

1. Build whole, **functionalist** architectures

Requirement:

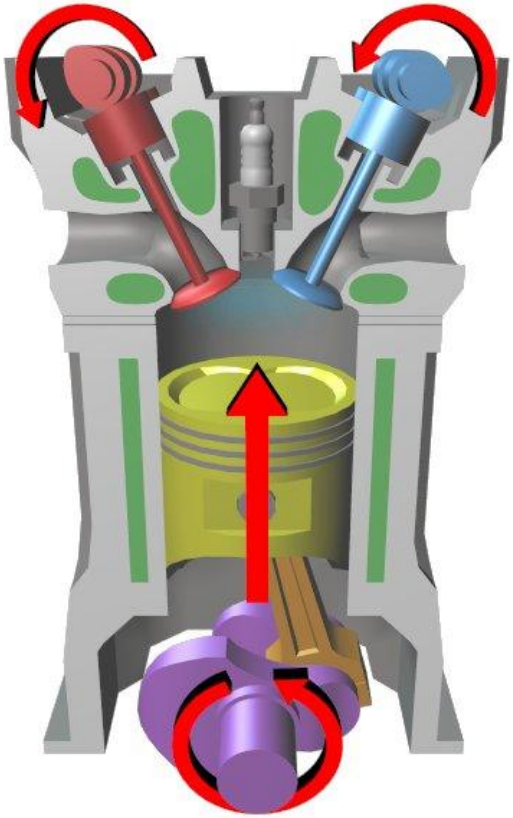
Dissection of system
into parts
and relationships
between them



#1: Build functionalist architectures

Requirement:

Dissection of system
into parts
and relationships
between them



Machines to explain the mind

Suppose there would be **a machine**, so arranged as **to bring forth thoughts, experiences and perceptions**; it would then certainly be possible to imagine it to be proportionally enlarged, in such a way as to allow entering it, like into a mill. This presupposed, **one will not find anything** upon its examination **besides individual parts, pushing each other**

(Gottfried Wilhelm Leibniz 1714)

Machines to explain the mind

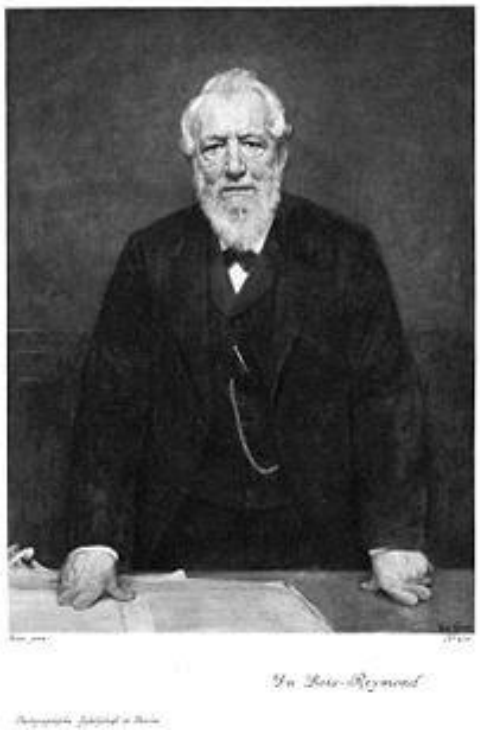
Perception, and what depends on it, is inexplicable in a mechanical way, that is, using figures and motions.

Suppose there would be **a machine**, so arranged as **to bring forth thoughts, experiences and perceptions**; it would then certainly be possible to imagine it to be proportionally enlarged, in such a way as to allow entering it, like into a mill. This presupposed, **one will not find anything** upon its examination **besides individual parts, pushing each other**—

and never anything by which a perception could be explained.

(Gottfried Wilhelm Leibniz 1714)

Emil du Bois-Reymond



Die astronomische Kenntnis des Gehirns, die höchste, die wir davon erlangen können, enthüllt uns darin nichts als bewegte Materie. Durch keine zu ersinnende Anordnung oder Bewegung materieller Teilchen aber läßt sich eine Brücke ins Reich des Bewußtseins schlagen.

Emil du Bois-Reymond, *Über die Grenzen des Naturerkennens (1872)*

LaMettrie's "L'Homme Machine" (1748)

- The world is entirely physical (a machine)
- Man is an animal, not much different from the great apes
- Man is a machine
- Whatever can be known about the soul can only be known through empirical scientific inquiry

Common ground?

Artificial Intelligence:

- The mind is a computational system
- The mind is **less** than a Turing machine
- Thinking, perception, feeling, volition, normativity, … have to be explained in terms of information processing
- We can test our ideas on how the mind works by running them as computer programs

Darthmouth Conference: 1956

John McCarthy

Nathan Rochester

Claude Shannon

Marvin Minsky

The explanatory gap

Descartes:

Res Cogitans vs. Res Extensa

Whoever counted the two substances?

The Leibnizians have created an unintelligible hypothesis. They have spiritualized matter, instead of materializing the spirit. How can one define something with a nature that is completely unknown to us?

Descartes and the Cartesians (…) have committed the same mistake. They have projected two different substances into man, just as if they had seen and counted them.

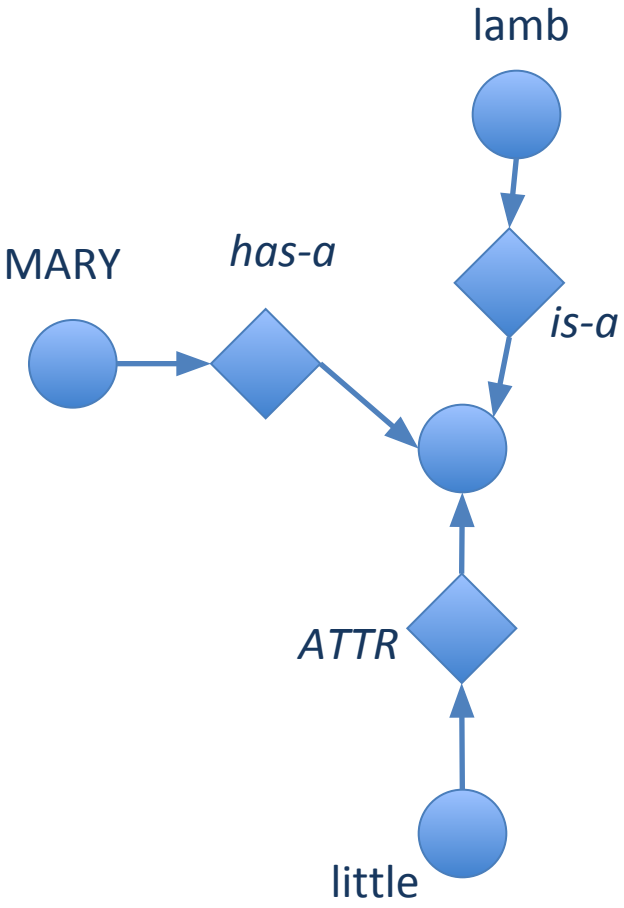
Rule-based systems (symbolic AI)

- regularity, systematicity, compositionality
- planning, inference, natural language processing

'MARY HAS-A LITTLE LAMB' :- has-a(MARY, lamb) AND little(lamb);

Rule-based systems

- regularity, systematicity, compositionality
- planning, inference, natural language processing



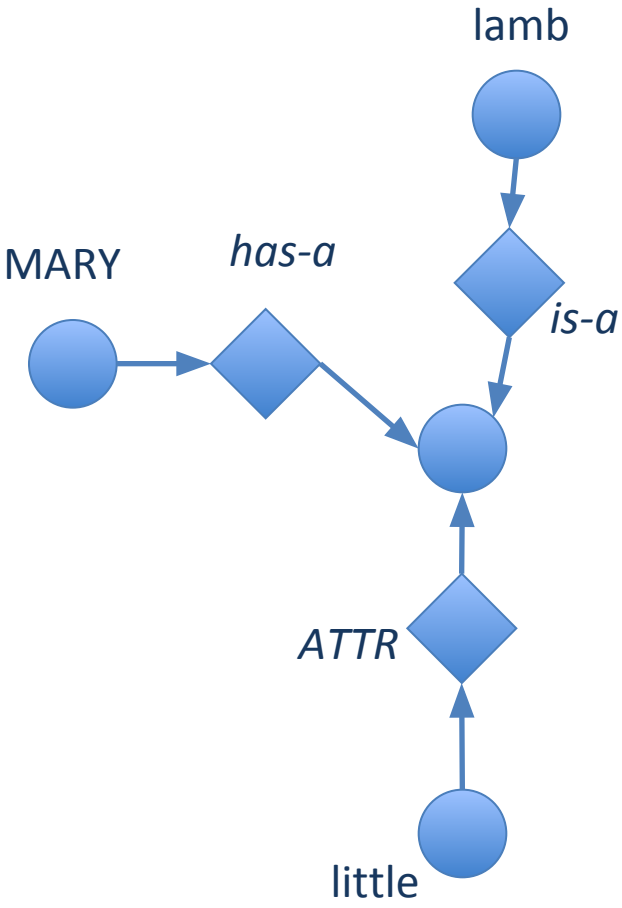
'MARY HAS-A LITTLE LAMB' :-
has-a(MARY, lamb) AND little(lamb);

Rule-based systems

Criticisms:

- *combinatorial explosion*

```
girl(MARY);
girl(X) :- human(X) AND female(X) AND little(X);
little(X) :- ATTR(smaller-than(X, most-of(Y|(is-a(Y,Z) AND is-a(X, Z)))));
female(X) :- (animal(X) OR plant(X)) AND has-a(X, female-reproductive-organs);
human(X) :- mammal(X) AND intelligent(X) AND NOT dolphin(X); ...
```

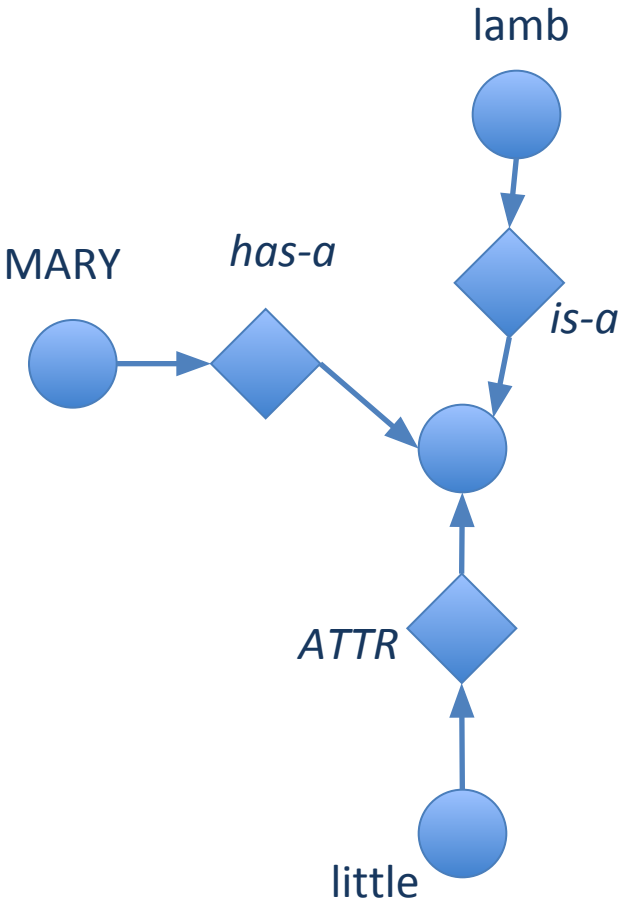


Rule-based systems

Criticisms:

- *combinatorial explosion*
- brittleness
- a-modality
- lack of grounding

→ problems of expert systems

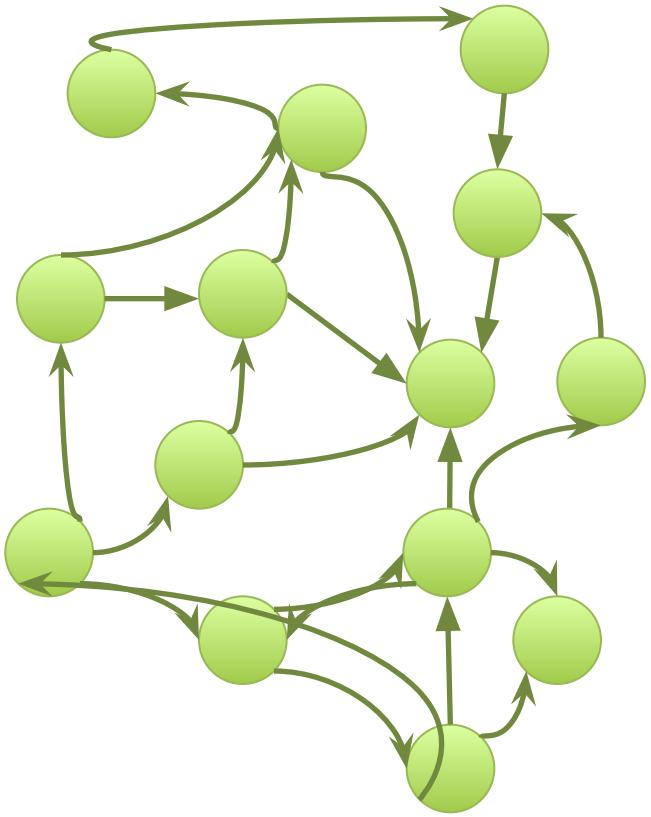


Connectionism

- distributed networks with spreading activation
- neural learning

easier adapt to

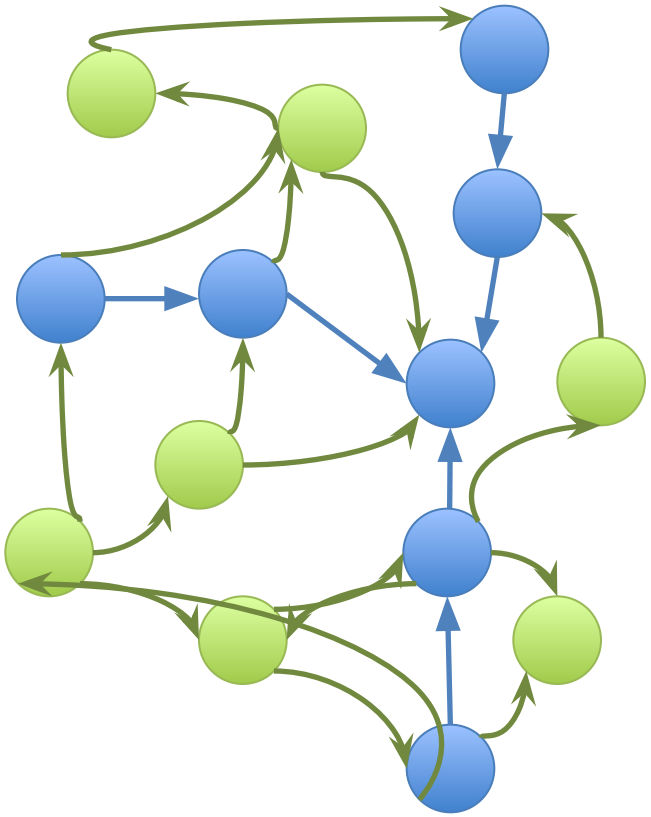
- pattern recognition
- low level sensor-motor coupling
- classification



Neuro-Symbolic Models

- Use methods from symbolic and connectionist systems in the same formalism

- symbolic links as a special case of connectionist links



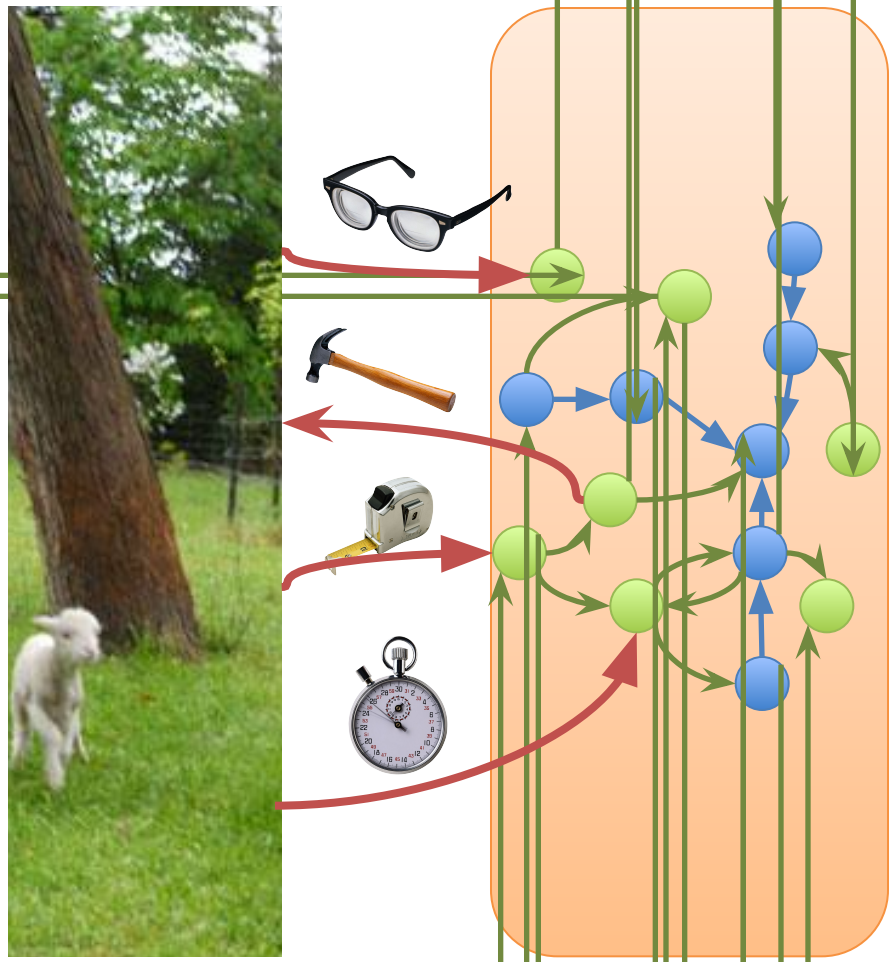
Deeper Criticisms

- social embedding?
- individual psychological development?
- rich physical environment?
- human sensory modalities?

→ *"New AI"*

New AI

- autonomous, learning agents, instead of inert problem solvers
- situated cognition
- rich, dynamic environment
- *embodiment* as an engineering solution to knowledge acquisition



The Extended Mind

one step beyond:

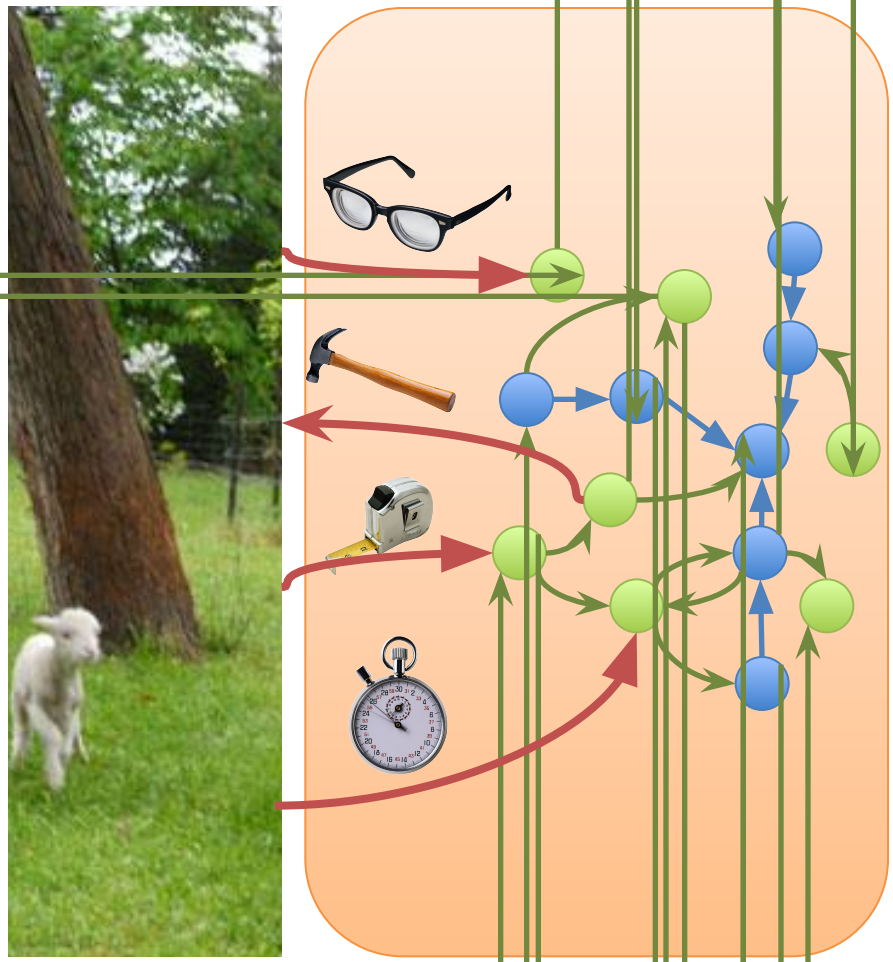
- tools, immediate environment should be considered part of mind

(Clark 2008)

no challenge to computational models

- add a slice of enviroment to the model

or add tool use capabilities

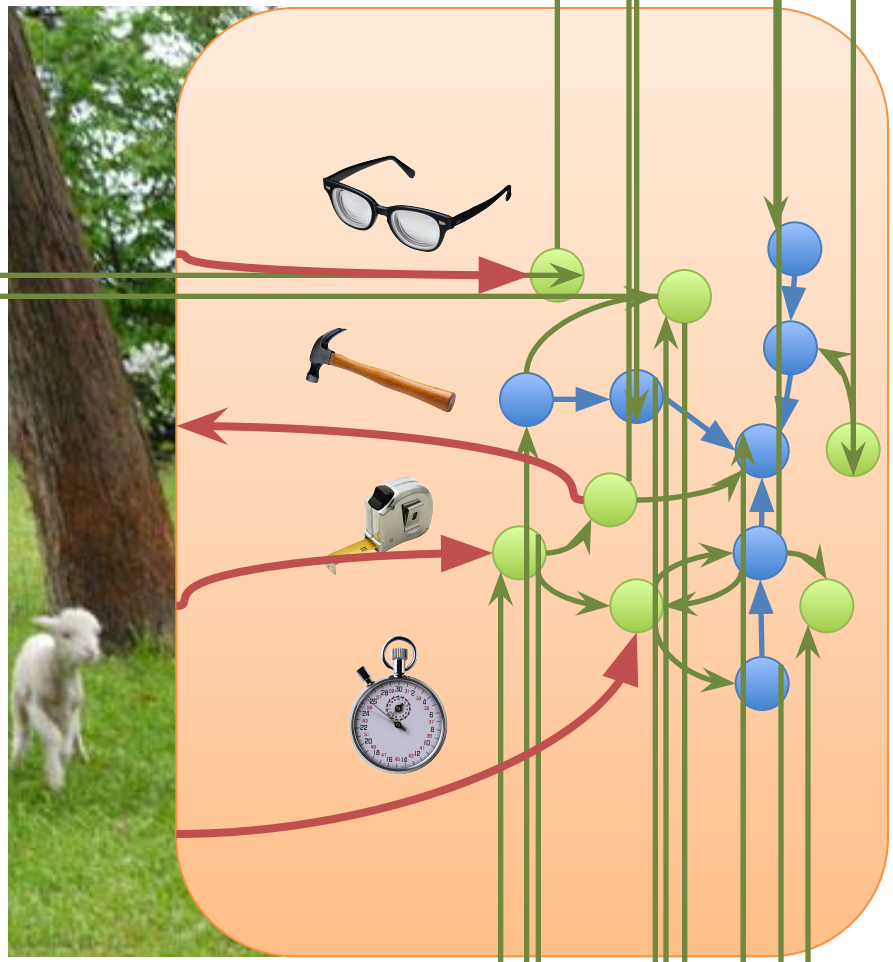


The Extended Mind

one step beyond:

- tools, immediate environment should be considered part of mind

(Clark 2008)



no challenge to computational models

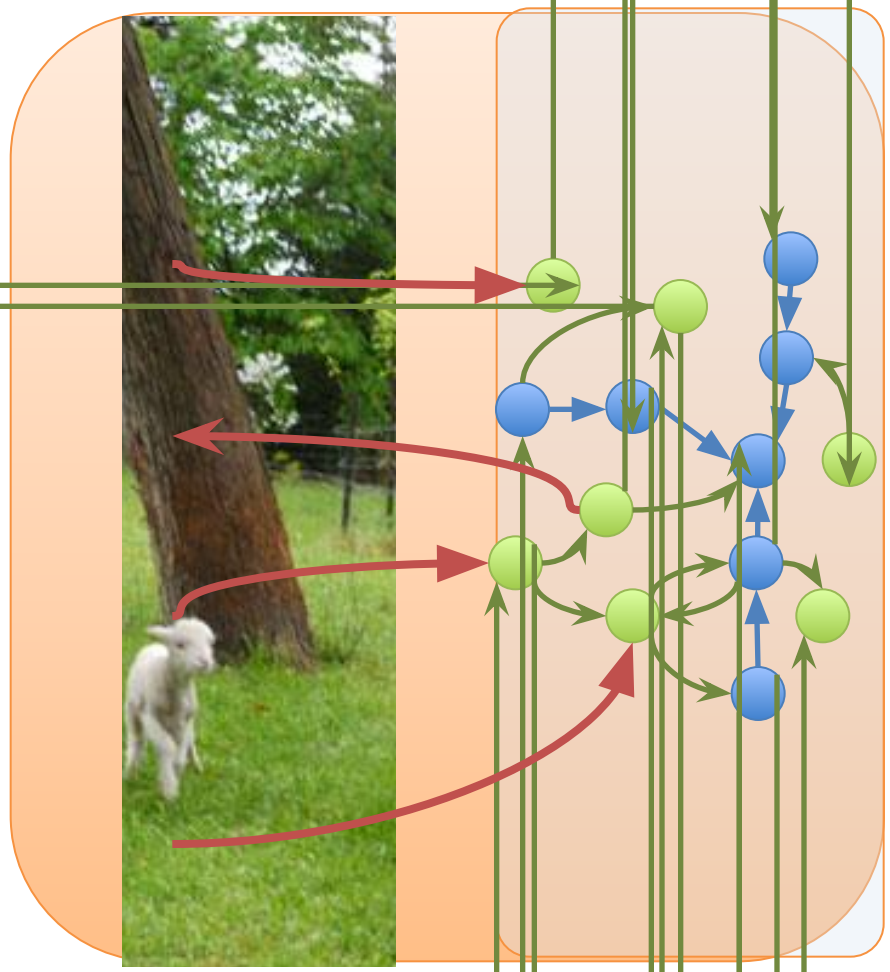
- add a slice of enviroment to the model

or add tool use capabilities

Enactivism

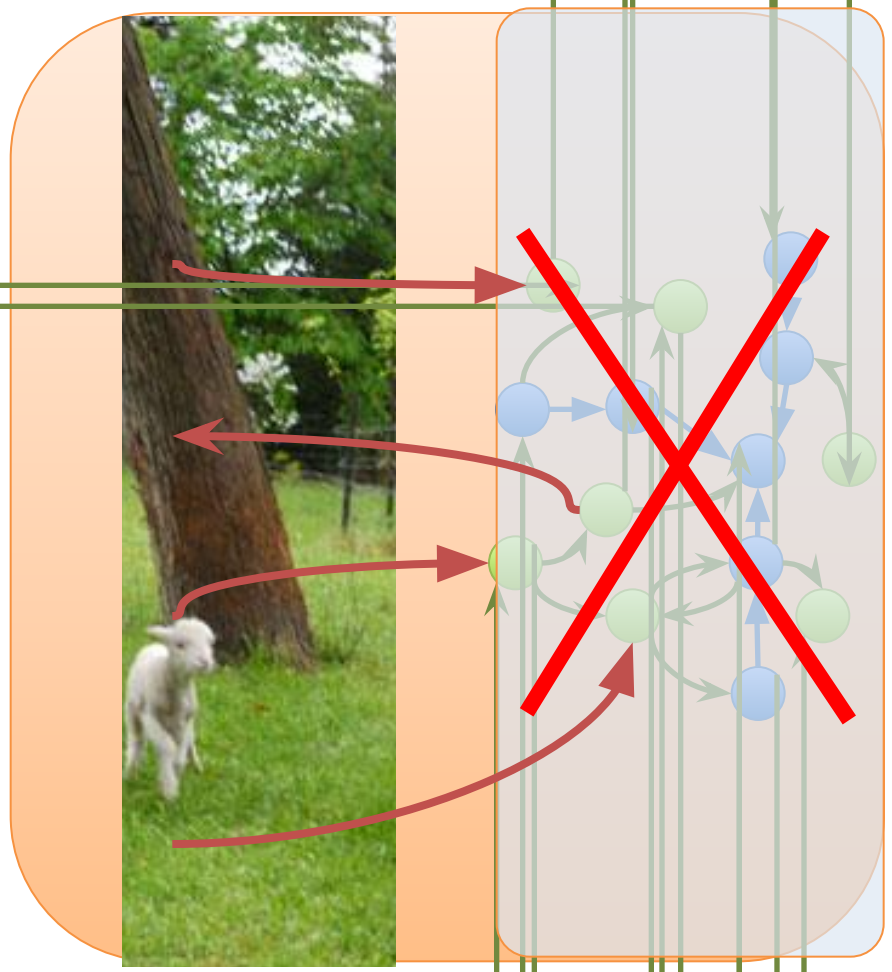
- cognition *is interaction* with the outside world

- ~~mind does *not* supervene~~ over activity of nervous system,
  but over interaction between body and world



Enactivism

- cognition *is interaction* with the outside world
- ~~mind does *not* supervene~~ over activity of nervous system,
  but over interaction between body and world
- Anti-Represenationalism



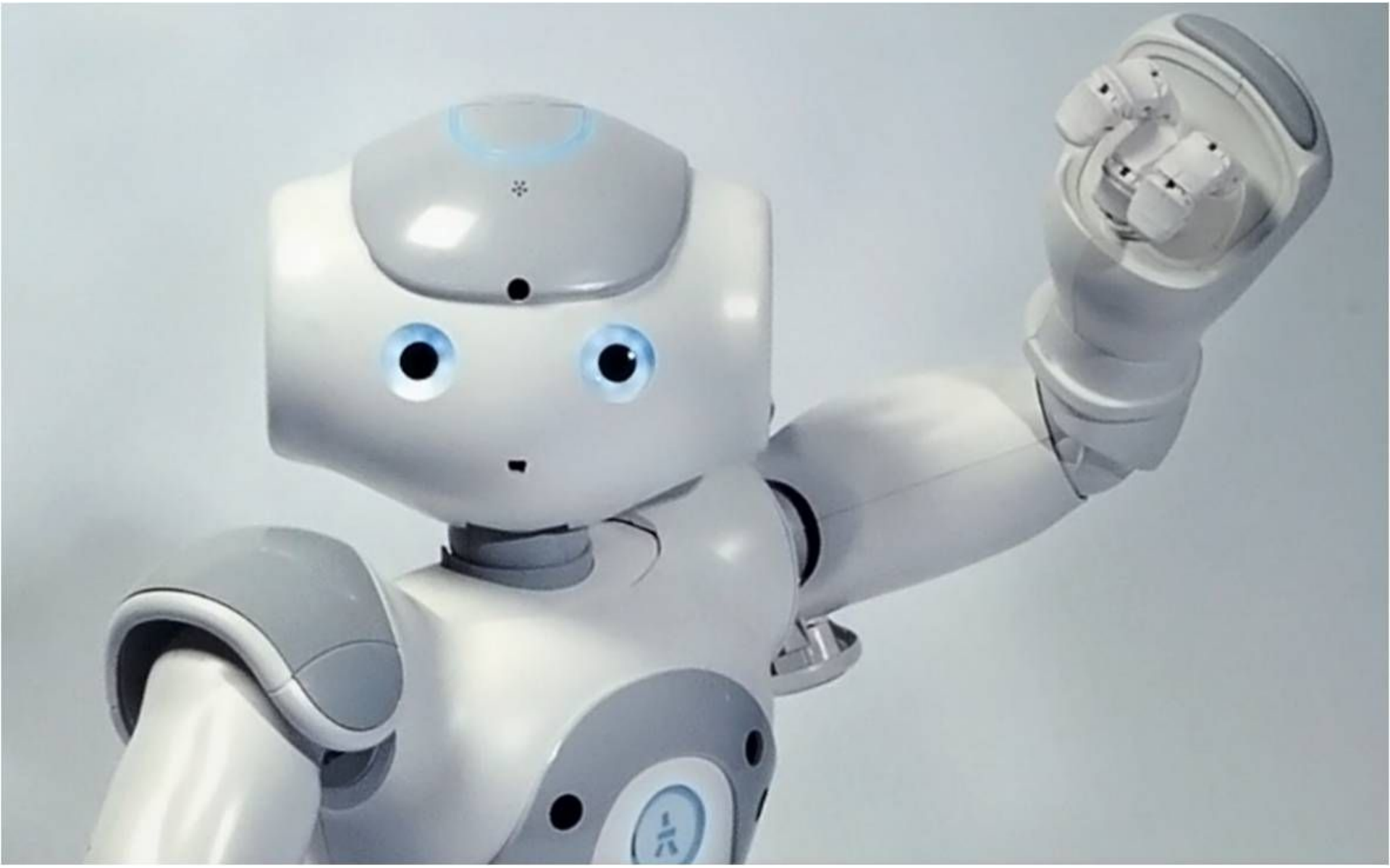
Enactivism and robotics

- attractive, because of call for bodily interaction!

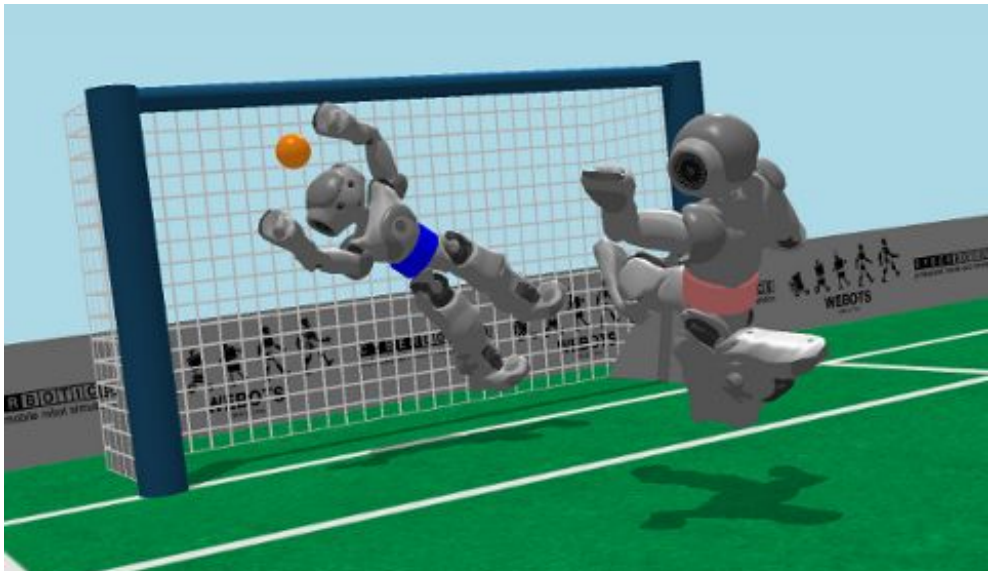
but: **enactivism and AI are incompatible**

- according to AI:
  - environment is information provider
  - cognition is information processing
- according to enactivism:
  - cognition takes place in environment
  - both cognition and environment are more than information processing

→ no AI system, computational model, or robot can ever perform cognition



Realism does not work for robots



- in AI and robotics, simulation and lab environment are in principle exchangeable
- for robot, no difference detectable,
  if input data have the same structural properties
- no similarity between internal program structure of simulator and simulation data necessary

→ no chance for the robot to discover the “real” structure of its world

- robot can only encode apparent regularities over data at its interface

Robot epistemology

- robot/computational agent can only have truth for *a-priori* and *synthetic* concepts (mathematics, including computational theories)
- for empirical contexts ⟶ only **suitability** of encoding

suitability:

- consistency
- completeness
- stability
- sparseness
- relevance
- maintenance and acquisition cost
- memory and access demands

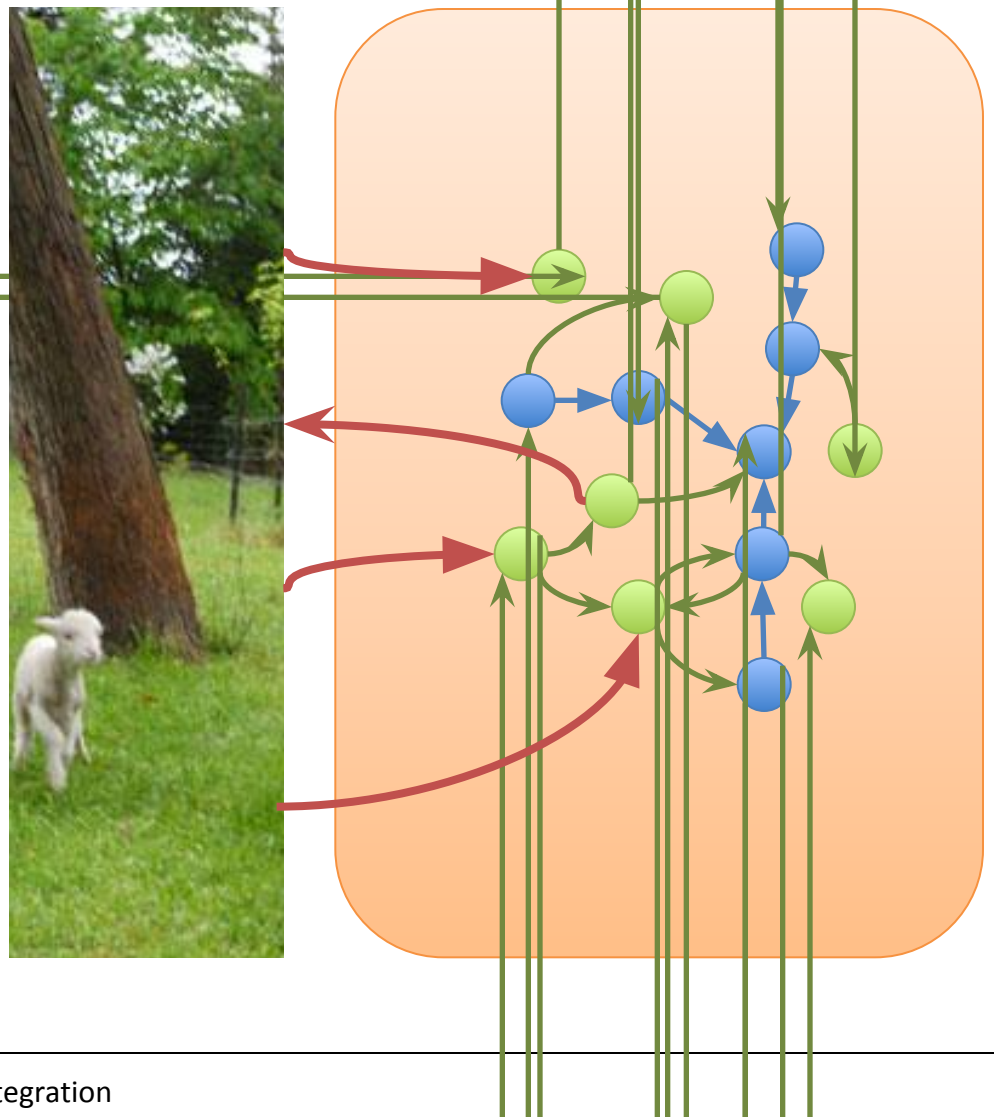
Epistemological Epoché

For human agents, epistemology is similar:

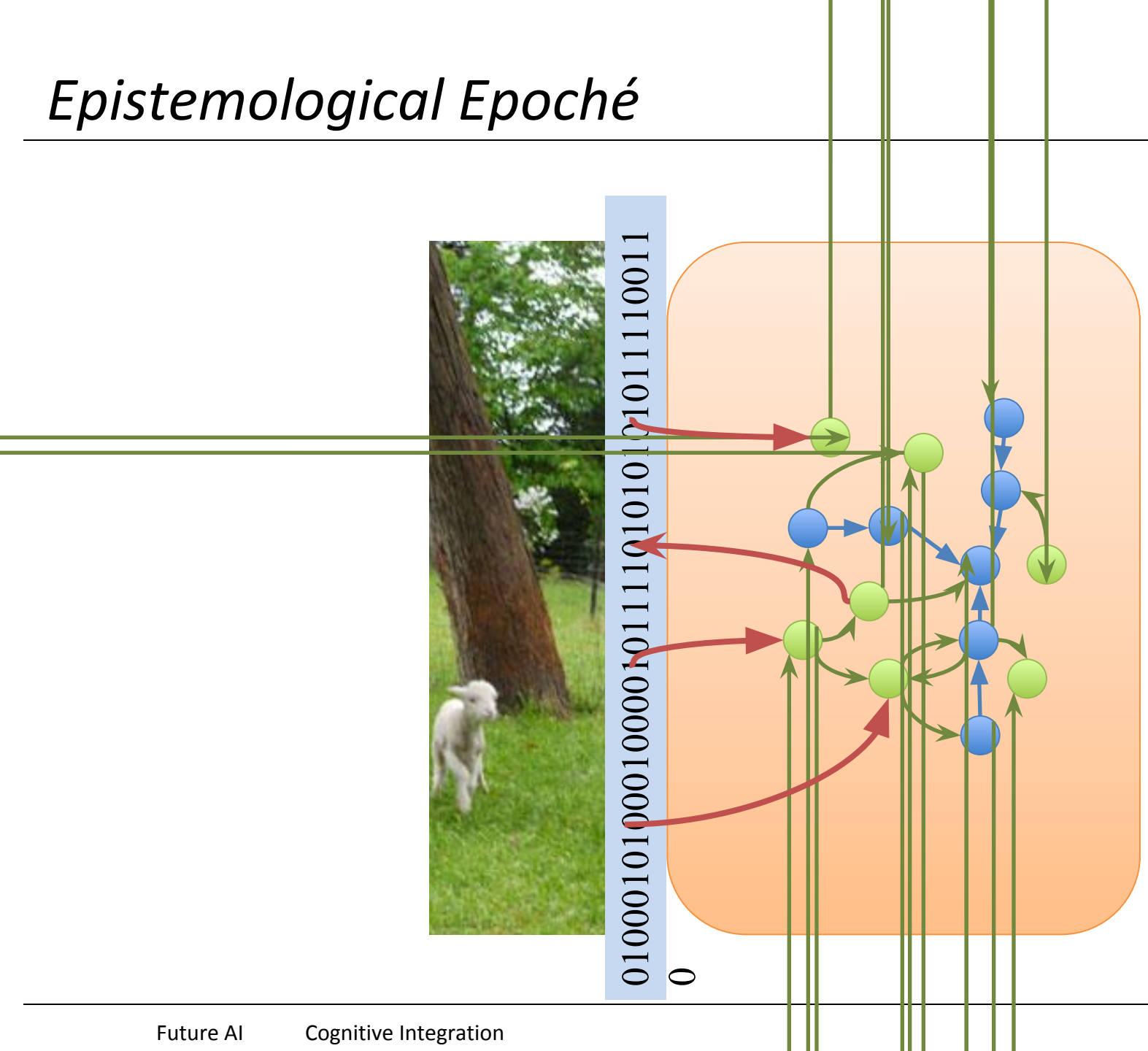
- *no direct access* to an external environment
- all access to the world is via a data transfer (electrical impulses through nerve endings)
- all sensory modalities, taken over lifetime, can be represented as *a finite vector of bits*
- all cognitive processing amounts to a re-encoding of this vector
- applies to every theory that admits interface between an information processing cognitive system and any kind of environment

→ epoché with respect to reality of objects of external, physical world

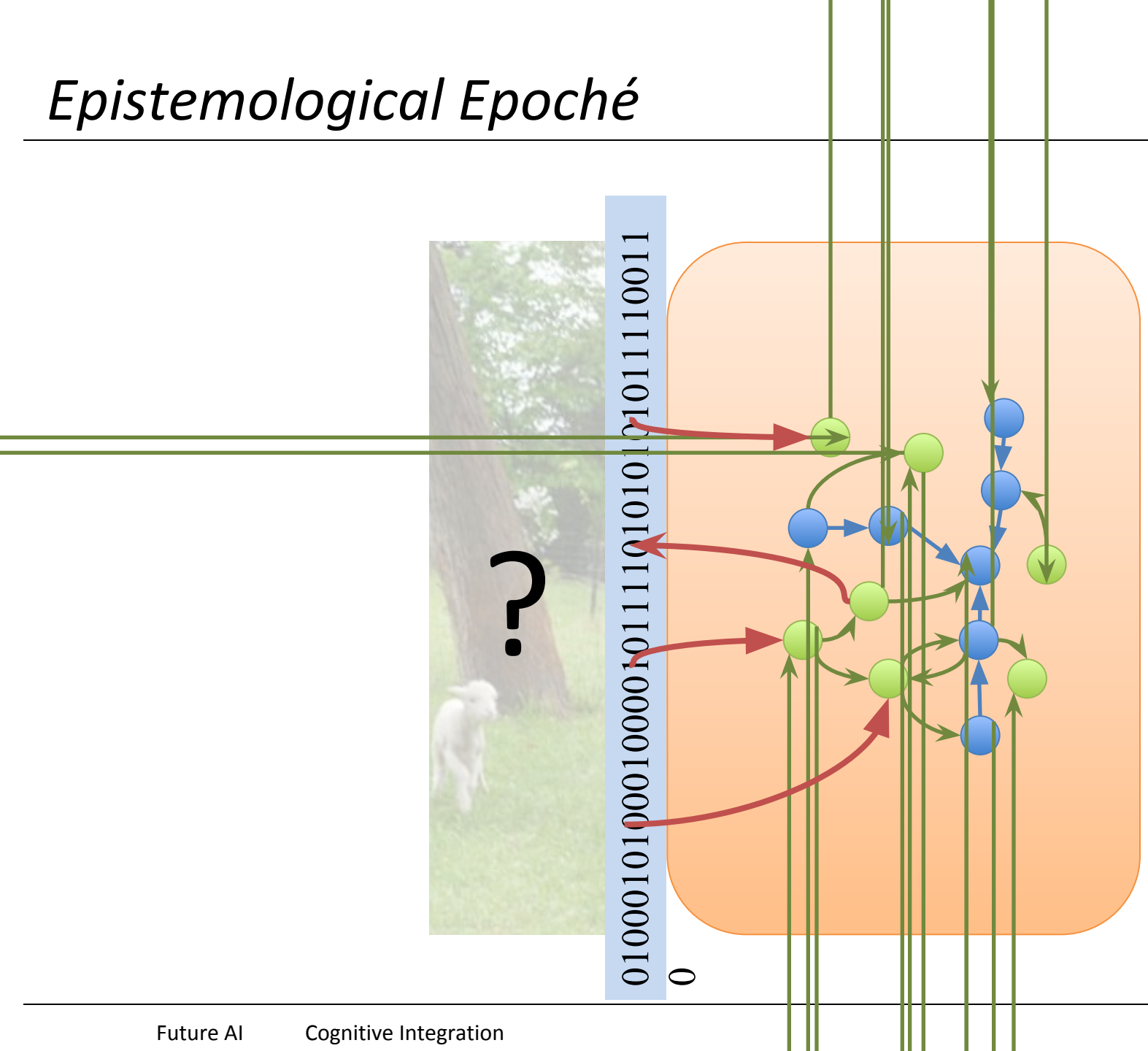
Epistemological Epoché



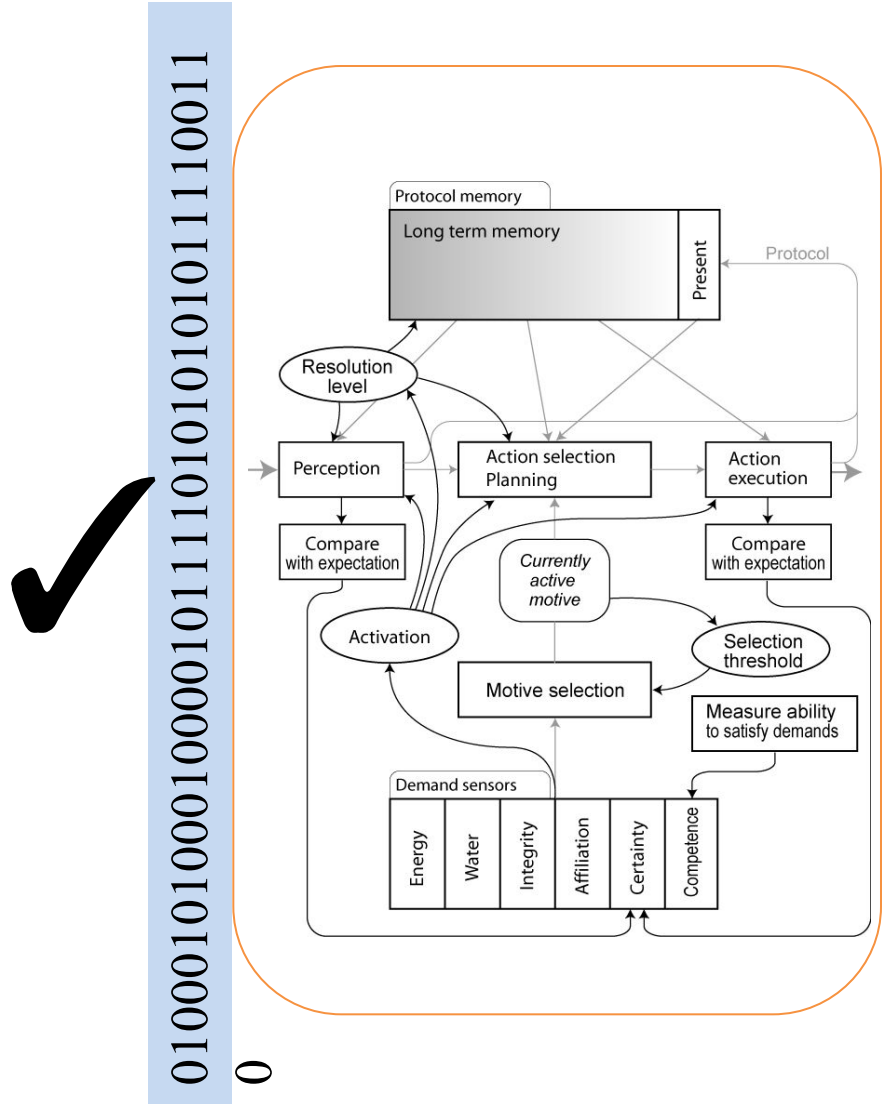
Epistemological Epoché



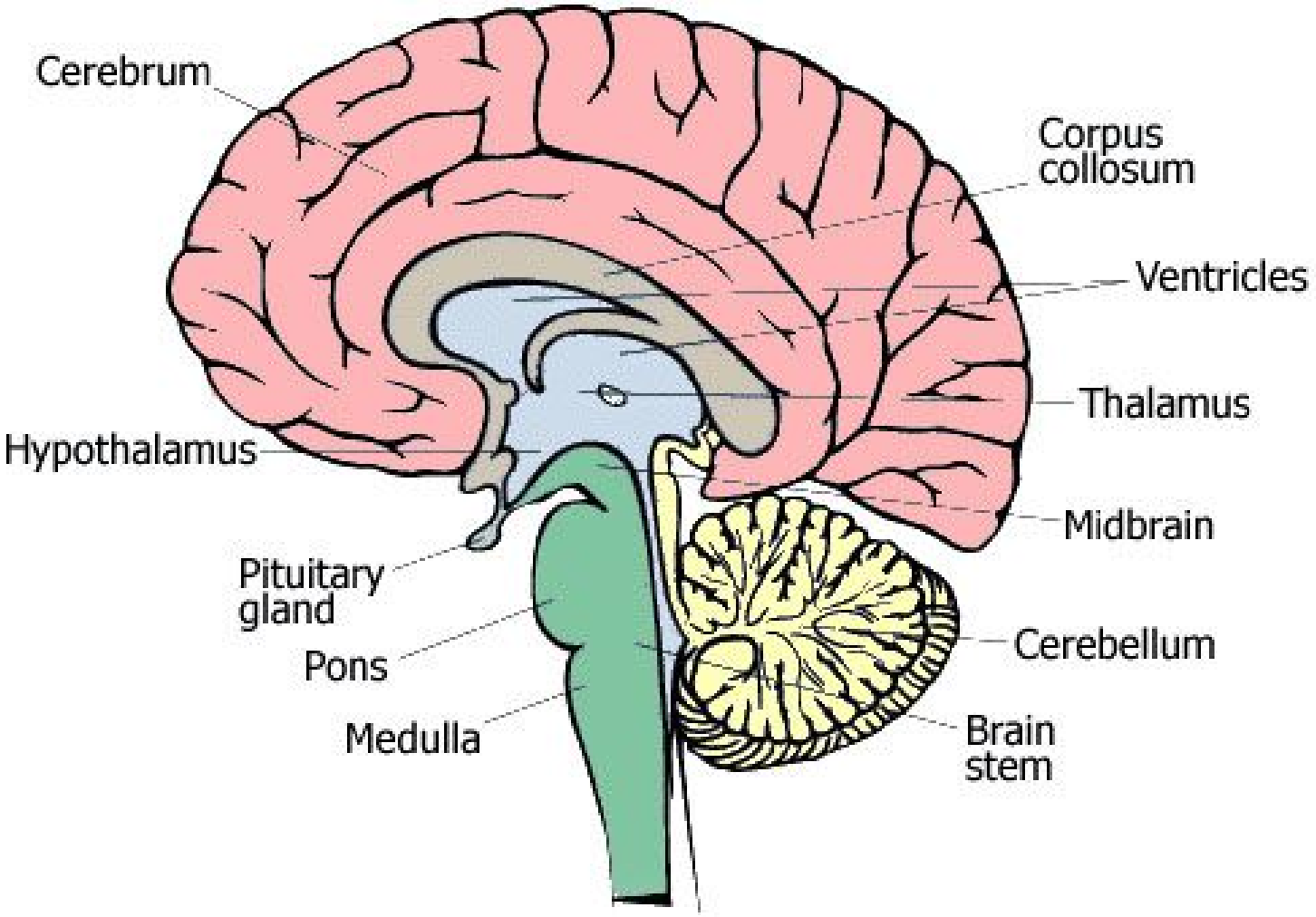
Epistemological Epoché



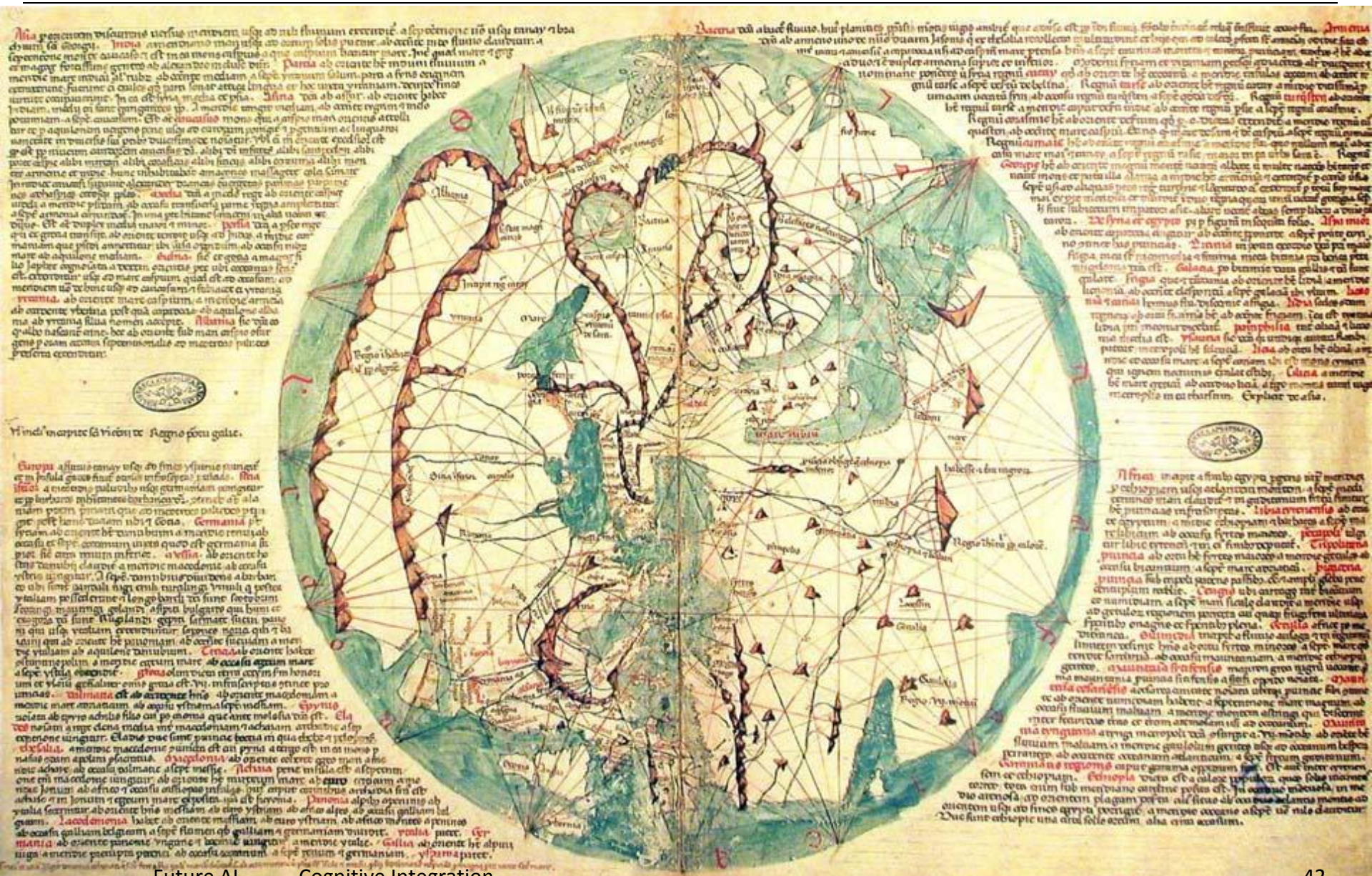
Epistemological Epoché



From brains to minds



Pietro Visconti: Mappa Mundi (1321)



Challenges for AI

- Re-integrate Cognitive Sciences
- Constructionist methodology: implementation to make complex theories testable
- Focus on broad models

- Science instead of engineering