CENTER FOR RESEARCH IN COGNITION & NEUROSCIENCES /CO3

UNIVERSITÉ LIBRE DE BRUXELLES

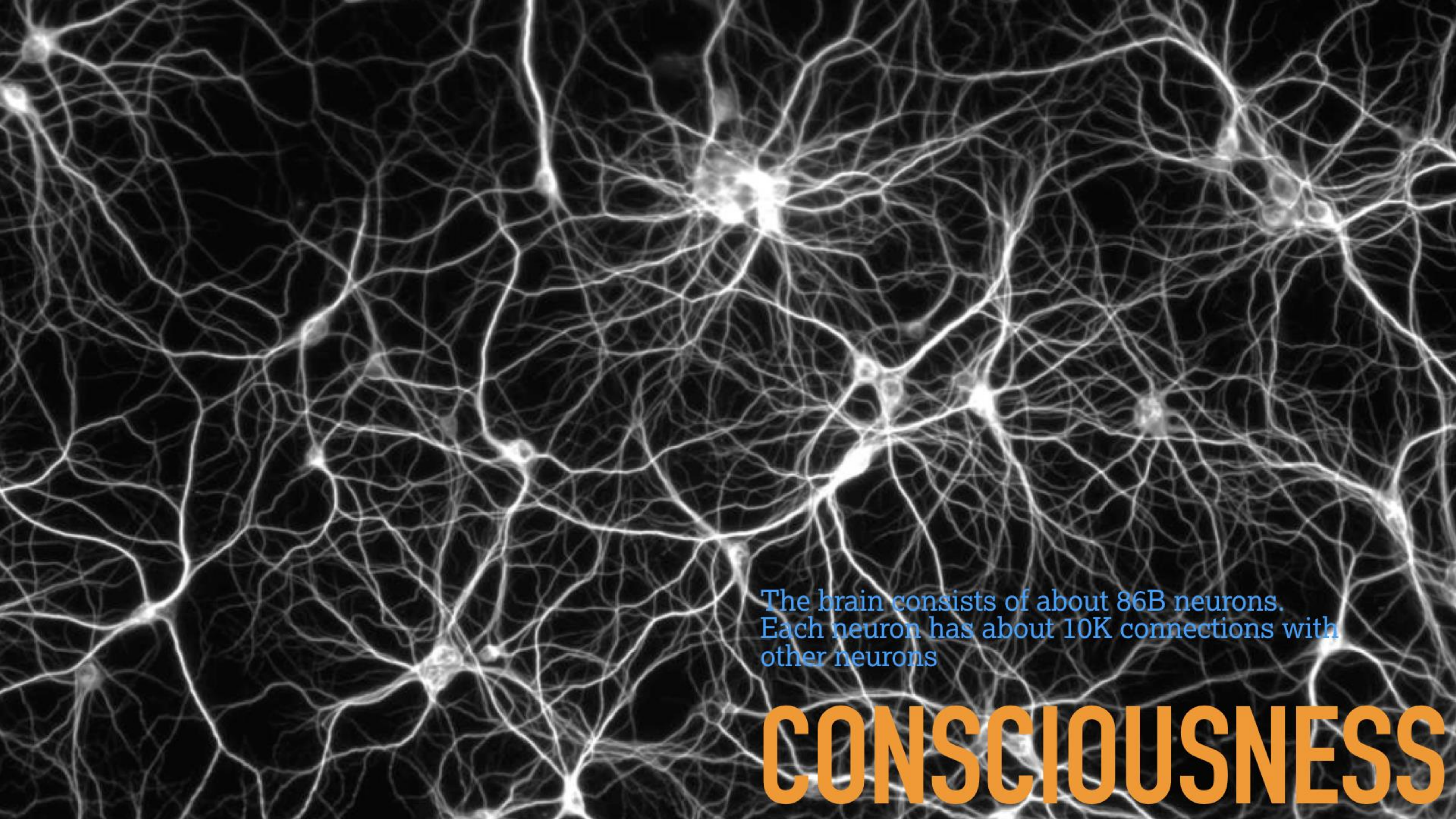
ULB

UNIVERSITÉ D'EUROPE

CONSCIOUSNESS

erc
European Research Council
Established by the European Commission

fondation ULB



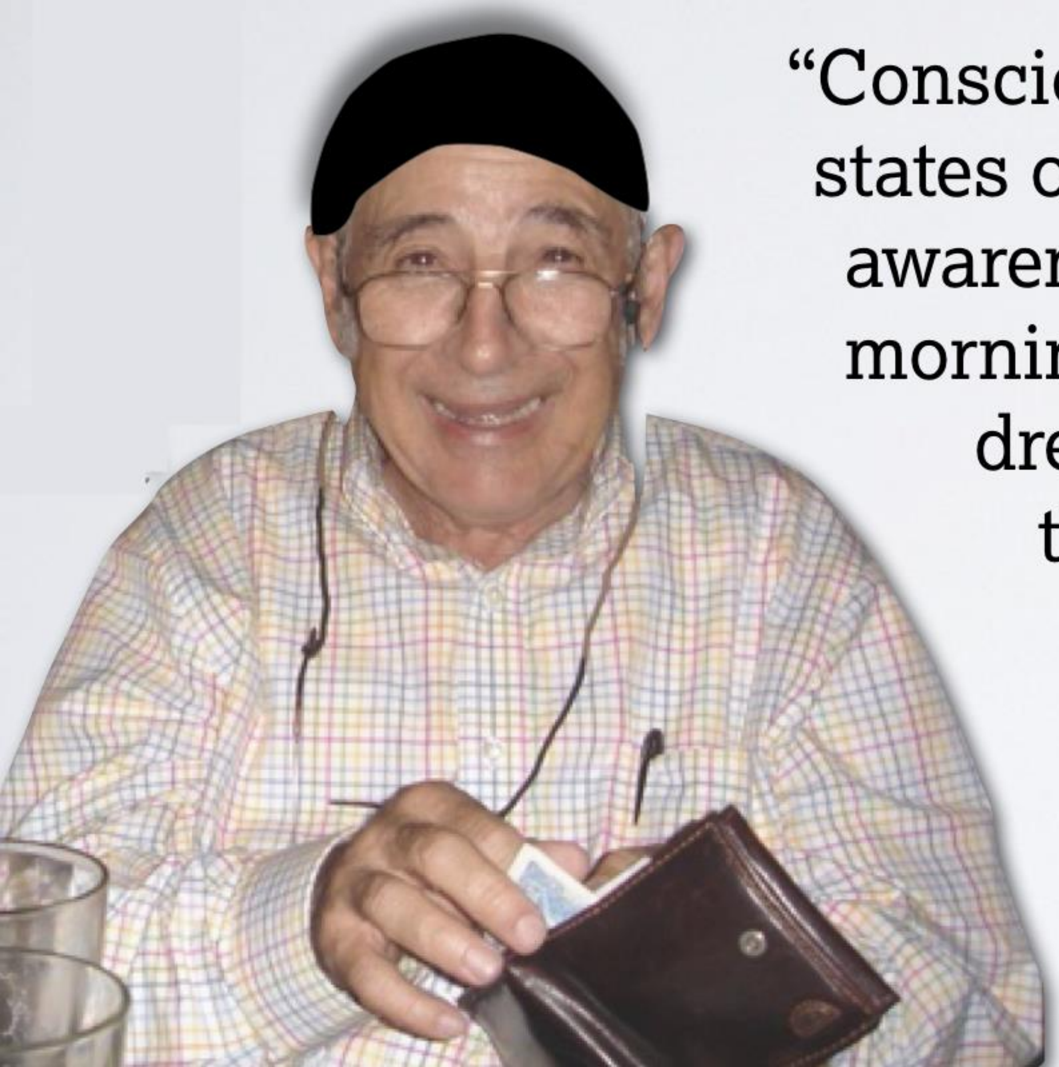
The brain consists of about 86B neurons. Each neuron has about 10K connections with other neurons

CONSCIOUSNESS

one of the most substantial scientific challenges of the 21st century is:

CONSCIOUSNESS

WHAT IS CONSCIOUSNESS?



“Consciousness consists of those states of sensation, or feeling, or awareness, which begin in the morning when we awake from a dreamless sleep and continue throughout the day until we fall into a coma, or die, or fall asleep again, or otherwise become unconscious”.

JOHN SEARLE

WHAT IS IT LIKE TO BE A BAT?

Nagel (1974): No matter how much we know about the brain of a bat, we'll never know what it feels like to chase insects at dusk...

WHAT IS CONSCIOUSNESS?

So are we much closer to grasping consciousness than when you started work on it, four decades ago?

Not very. I
ust scientific –
knowing what
ious and so forth
– but philosoph
e phenomenon of
consciousness. T
f how conscious
experience eme
, say, the redness
of red or fe
hat scientific
questions to a
s to ask because
t.

MARGARET BODEN

WHAT IS CONSCIOUSNESS?

So are we much closer to grasping consciousness than when you started work on it, four decades ago?

Not very. I think the fundamental problems aren't just scientific – knowing what's going on in the brain when we're conscious and so forth – but philosophical questions, and in particular about the phenomenon of consciousness. This concerns the so-called hard problem of how conscious experience emerges from matter, and why we experience, say, the redness of red or feel pain. It isn't just that we're not sure what scientific questions to ask; it's that we don't know what questions to ask because we don't know what we're talking about.

MARGARET BODEN

THE HARD PROBLEM

“Numerous books and have appeared recently,
and one might think ality, these works have
ignored the hard pro ne might call the ‘easy
problems’ of consciou information? How does
it integrate informa reports on our mental
states? These questio them does not solve the
hard problem: Why i ssing is accompanied by
subjective experience.

DAVID CHALMERS

THE HARD PROBLEM

“Numerous books and articles dedicated to consciousness have appeared recently, and one might think that there is progress. But in reality, these works have ignored the hard problem. Often, they concern what one might call the ‘easy problems’ of consciousness: How does the brain process information? How does it integrate information? How do we produce verbal reports on our mental states? These questions are interesting, but answering them does not solve the hard problem: Why is it the case that information processing is accompanied by subjective experience?”

DAVID CHALMERS

THE EASY PROBLEMS

All of this goes on without awareness — *Consciousness is not intelligence*

THE EASY PROBLEMS

All of this goes on without awareness — Consciousness is not sensitivity

THE SUBJECT

"It seems absurd to us that a pain, a mood, a wish should rove about the world without a bearer, independently. An experience is impossible without an experient. The inner world presupposes the person whose inner world it is"

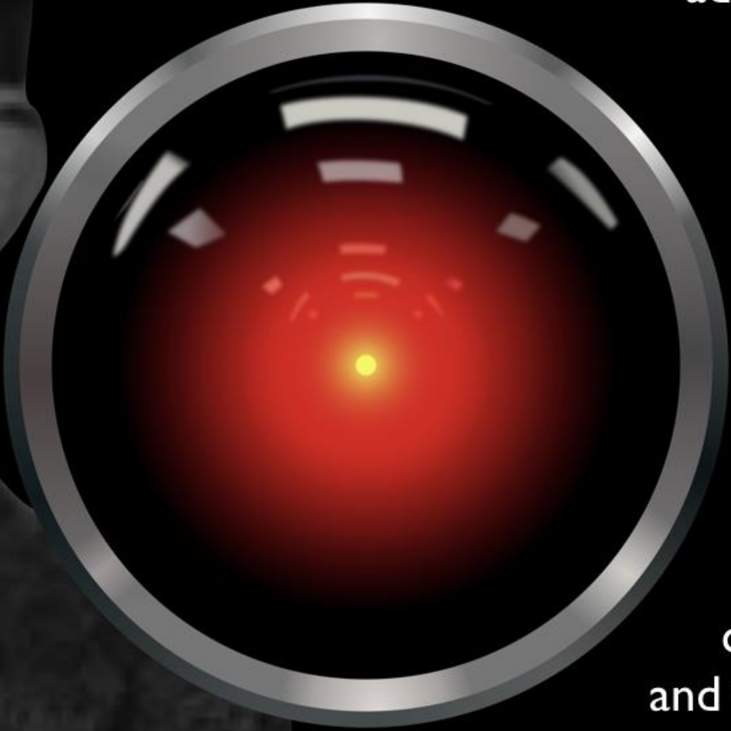
GOTTLOB FREGE

MACHINE CONSCIOUSNESS?

Alan Turing (1950)

On the “argument from consciousness”: This argument is very well expressed in Professor Jefferson's Lister Oration for 1949, from which I quote. “Not until a machine can write a sonnet or compose a concerto because of thoughts and emotions felt, and not by the chance fall of symbols, could we agree that machine equals brain-that is, not only write it but know that it had written it. No mechanism could feel (and not merely artificially signal, an easy contrivance) pleasure at its successes, grief when its valves fuse, be warmed by flattery, be made miserable by its mistakes, be charmed by sex, be angry or depressed when it cannot get what it wants.”

In short then, I think that most of those who support the argument from consciousness could be persuaded to abandon it rather than be forced into the solipsist position They will then probably be willing to accept our test.



John Searle (2001)

“I will argue that in the literal sense the programmed computer understands what the car and the adding machine understand, namely, exactly nothing.”

John Searle (2004)

“The fact that brain processes cause consciousness does not imply that only brains can be conscious. The brain is a biological machine, and we might build an artificial machine that was conscious; just as the heart is a machine, and we have built artificial hearts. Because we do not know exactly how the brain does it we are not yet in a position to know how to do it artificially.”(Biological Naturalism)

Is the thermostat conscious of temperature?

No. The thermostat is *sensitive* to temperature, but it is not *conscious* of temperature.

Why not?

Because the thermostat does not know *that* it is sensitive to temperature. It just is sensitive to temperature.

Conscious knowledge, however, is knowledge that you *know* you possess.

It would be easy to make it so that the thermostat is able to report on its internal states.

But that would be faking it.

Why?

Because we know that the thermostat does not *care* about temperature. It doesn’t even care about its own existence! It doesn’t have experiences because *nothing ever means anything to it.*

What would it take for the thermostat to care about its own states?

It would take the ability for the thermostat to be sensitive to its environment and to its own states *in a way that matters to it.*

It would take the ability to have goals, to pursue them, to avoid danger, to fall in love, to worry about one's own existence, and so on.

That is the machinery of agenthood.

What does it take for something to turn into an agent?

It takes a lot of different things, but all these things require the ability to *learn*.

Why is that?

Because learning is necessary to *grow a self :* to know what one wants, to develop preferences, to seek rewarding states, to learn about good and bad things, &c.

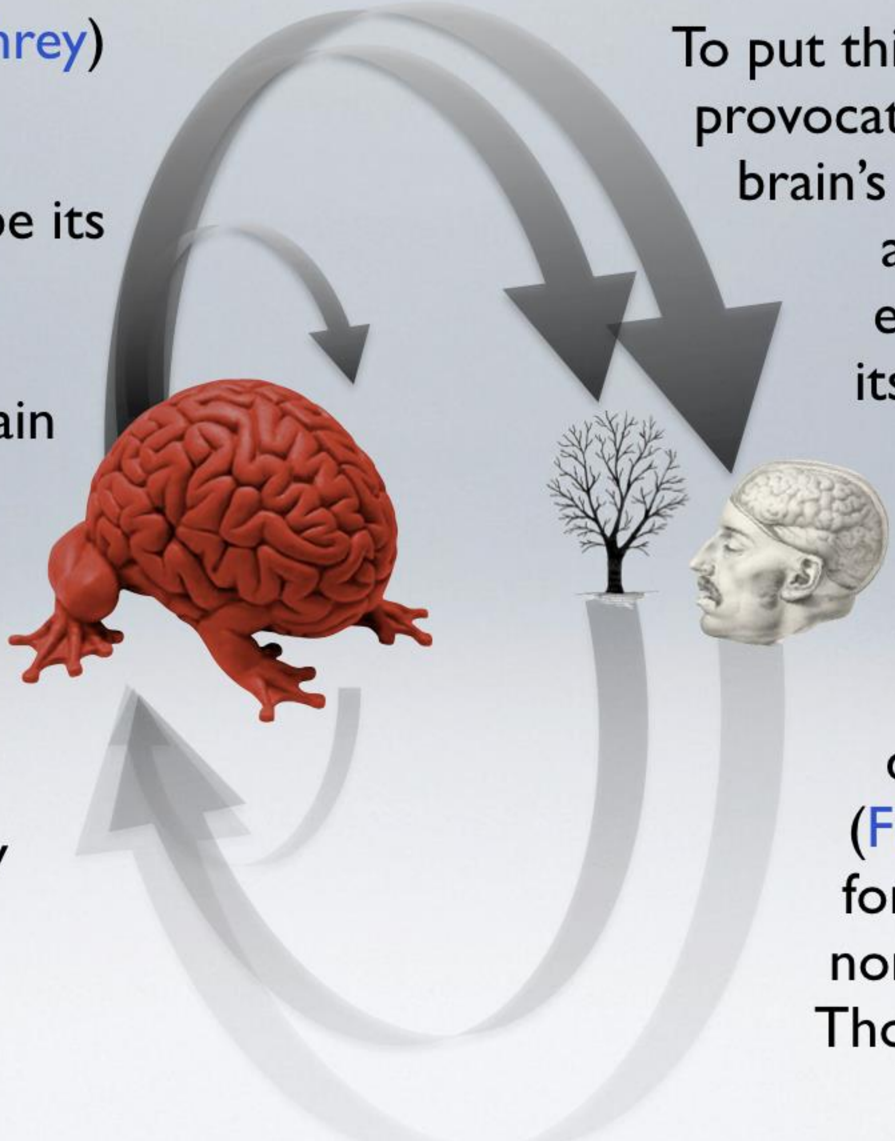
Thus, awareness requires agenthood

THE RADICAL PLASTICITY THESIS

In the beginning is action (Humphrey)

The brain continuously and unconsciously learns to redescribe its own activity to itself by assessing (Clark & Karmiloff-Smith) the consequences of action in the brain itself (the inner loop), on behaviour (the action loop), and on the behaviour of others (the mind loop).

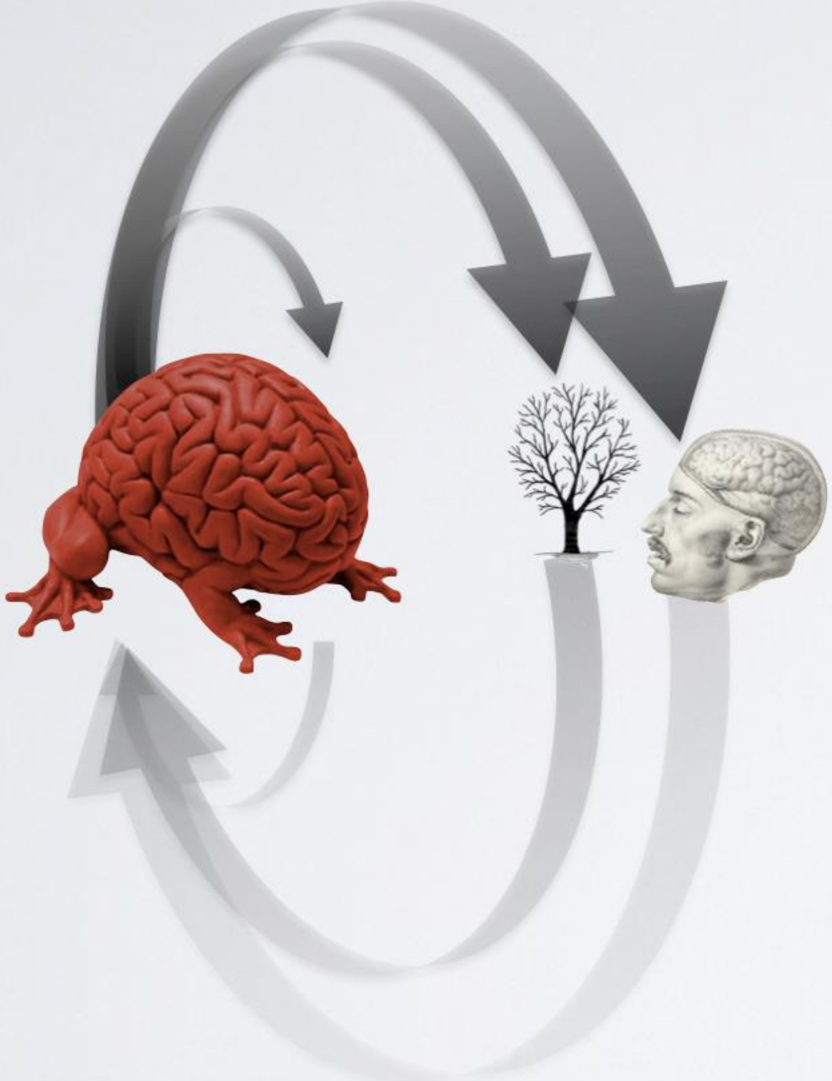
The three loops depend on each other, forming a tangled hierarchy (Hofstadter's strange loop).



To put this claim even more provocatively: Consciousness is the brain's (non-conceptual) theory about itself, gained through experience interacting with itself, with the world and with other people (Frith)

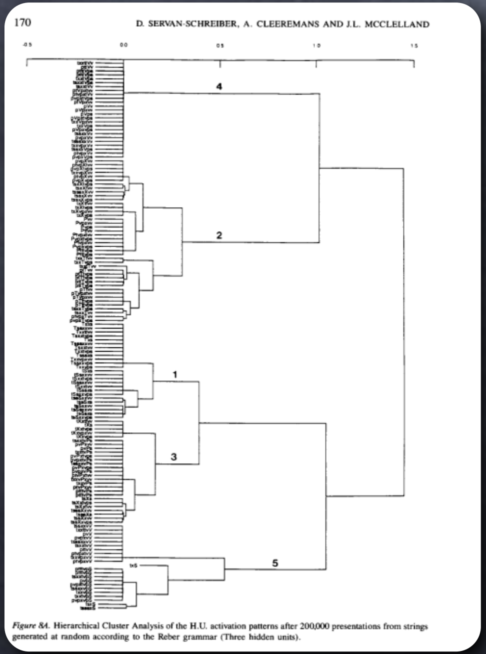
Consciousness depends on the operation of unconscious prediction-driven learning mechanisms (Friston's predictive coding) — a form of enactive (O'Regan), non-conceptual Higher-Order Thought Theory (Rosenthal)

ONE SYSTEM LOOKING AT (PART OF) ITSELF?



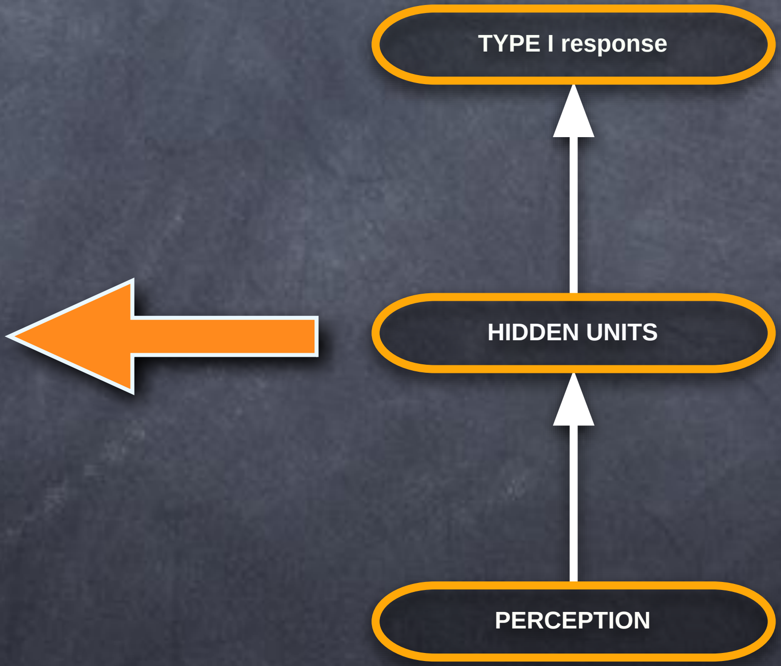
- Higher-Order Thoughts: requires HOTs, or representations about representations
- Metacognition: Requires one system judging the performance of another
- Predictive Processing: Requires internal models, or minimally one system making predictions about another system

REPRESENTATIONAL REDESCRIPTION



170 D. SERVAN-SCHREIBER, A. CLEEREMANS AND J.L. MCCLELLAND

Figure 8A. Hierarchical Cluster Analysis of the H.U. activation patterns after 200,000 presentations from strings generated at random according to the Reber grammar (Three hidden units).



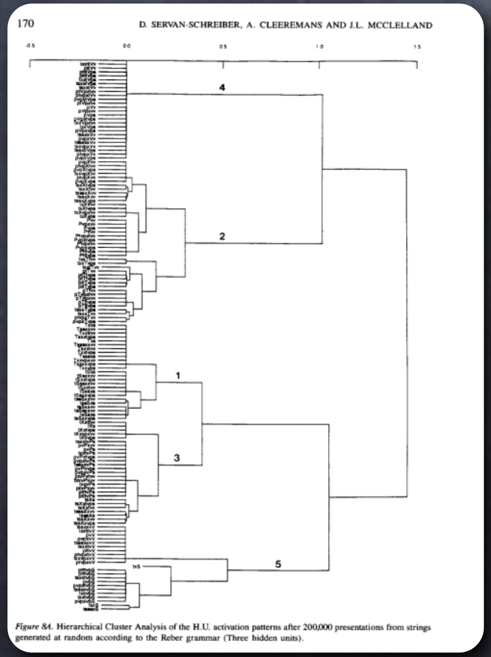
TYPE I response

HIDDEN UNITS

PERCEPTION

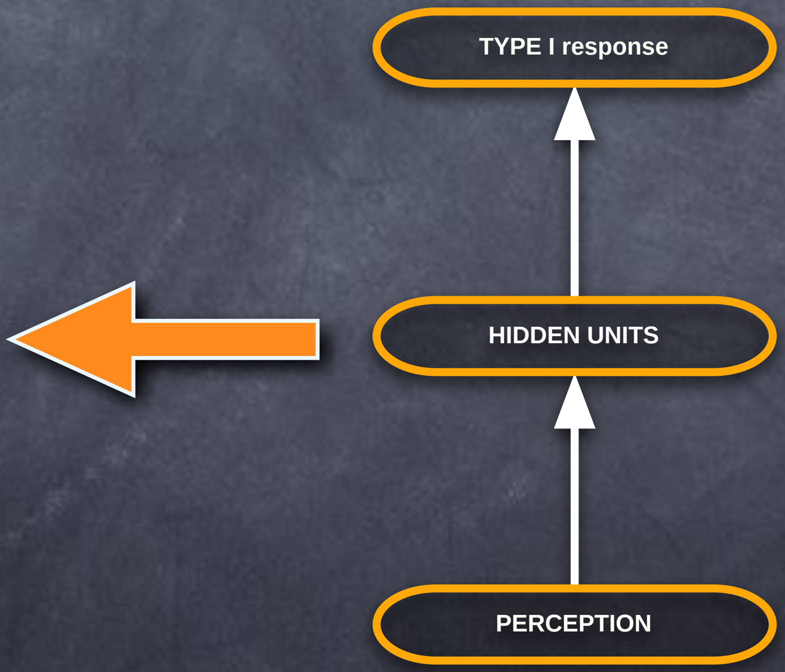
REPRESENTATIONAL REDESCRIPTION

The knowledge is forever embedded in the causal chain implemented by the network. All the network can do is project this knowledge onto action. It is knowledge “in the network” vs. knowledge “for the network” (Clark & Karmiloff-Smith)



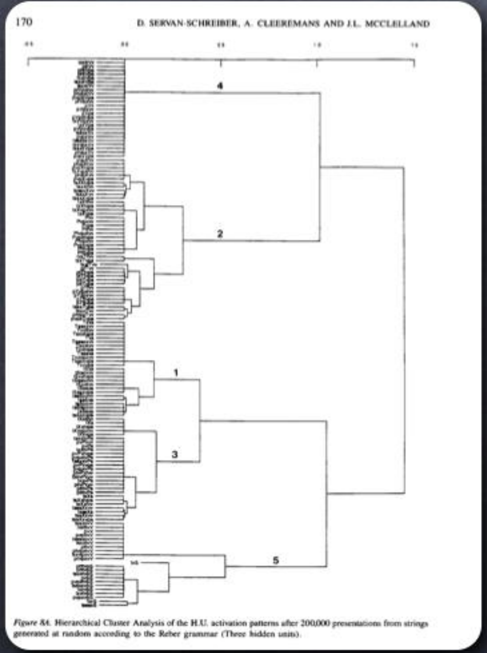
170 D. SERVAN-SCHREIBER, A. CLEEREMANS AND J.L. MCCLELLAND

Figure 8A. Hierarchical Cluster Analysis of the H.U. activation patterns after 200,000 presentations from strings generated at random according to the Reber grammar (Three hidden units).



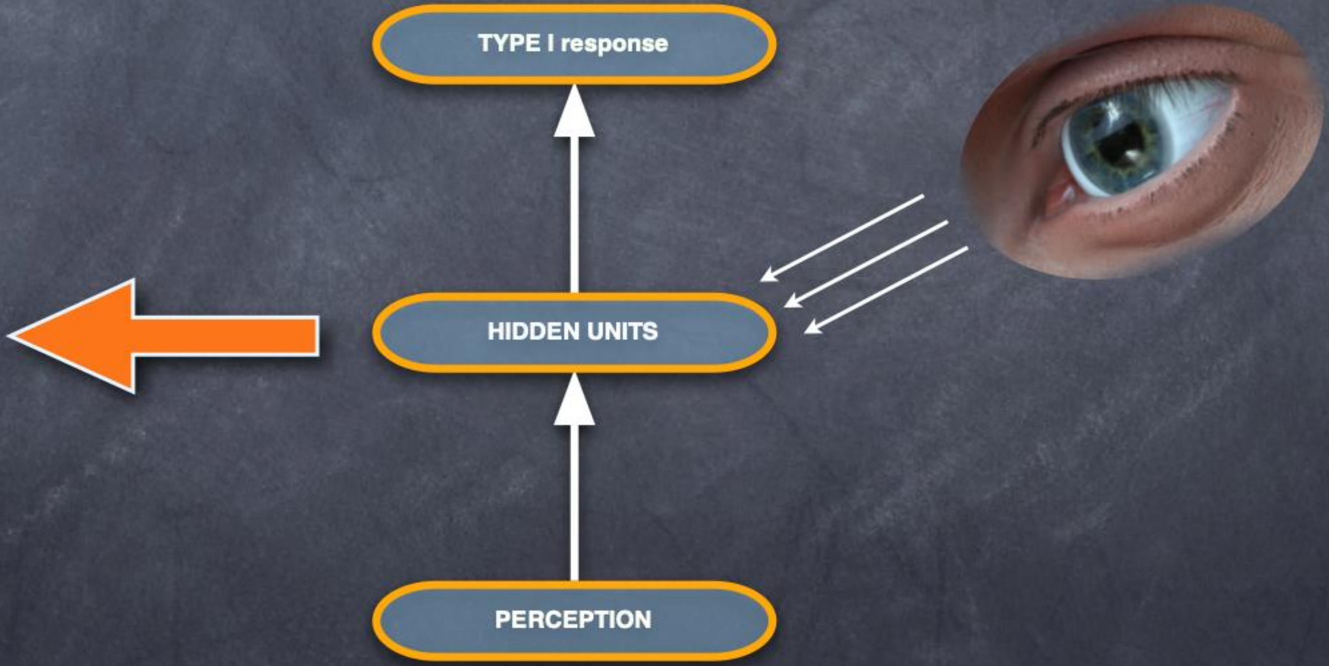
REPRESENTATIONAL REDESCRIPTION

The knowledge is forever embedded in the causal chain implemented by the network. All the network can do is project this knowledge onto action. It is knowledge "in the network" vs. knowledge "for the network" (Clark & Karmiloff-Smith)



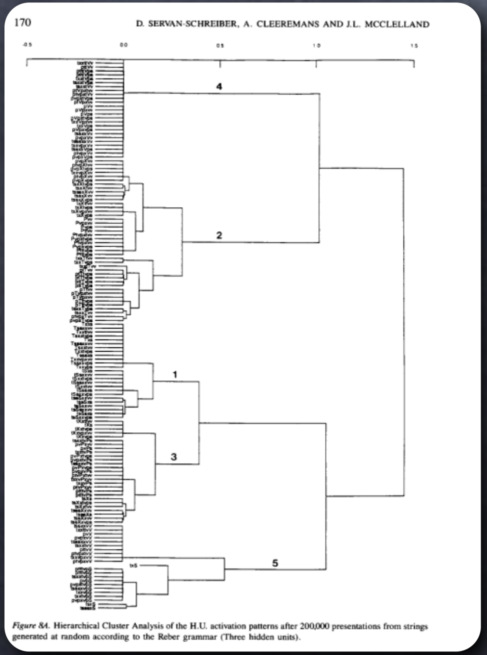
170 D. SERVAN-SCHREIBER, A. CLEEREMANS AND J.L. MCCLELLAND

Figure 8A. Hierarchical Cluster Analysis of the H.U. activation patterns after 200,000 presentations from strings generated at random according to the Reber grammar (Three hidden units).



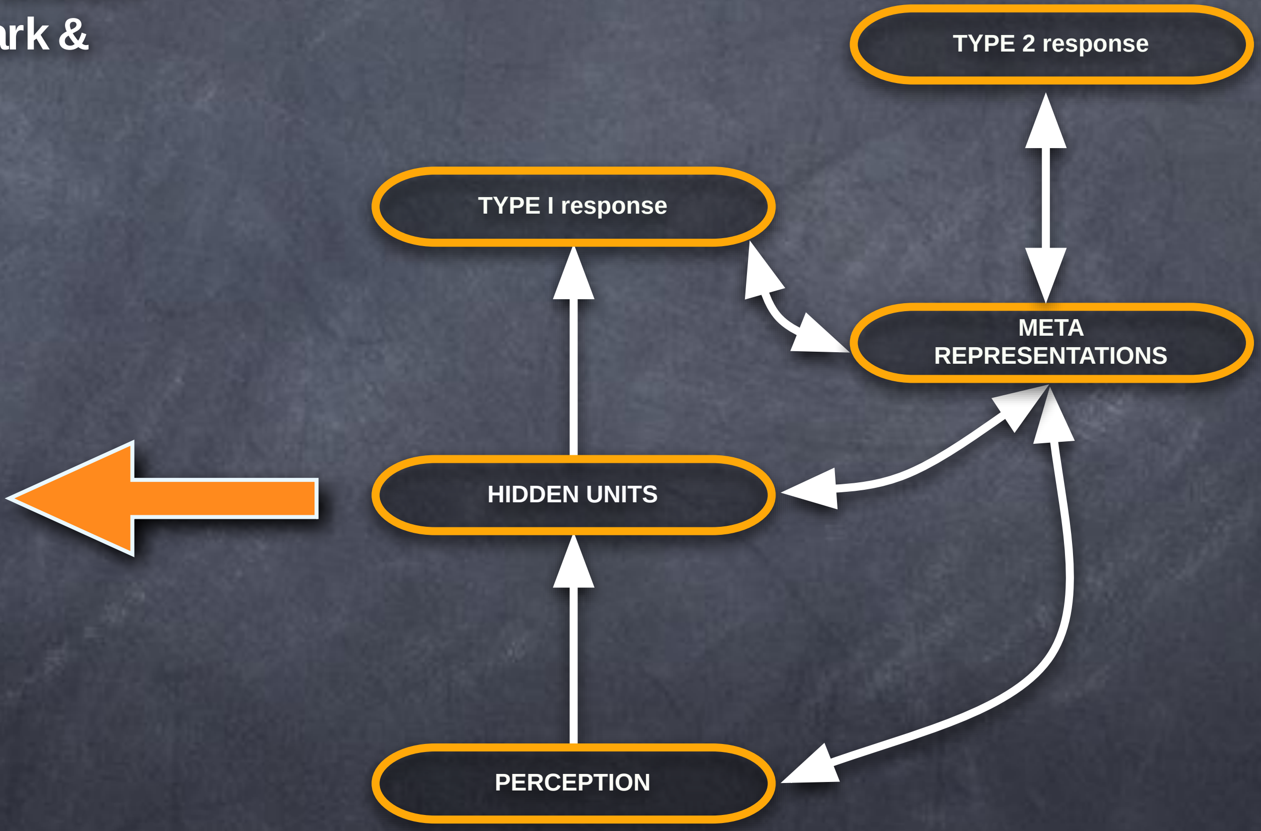
REPRESENTATIONAL REDESCRIPTION

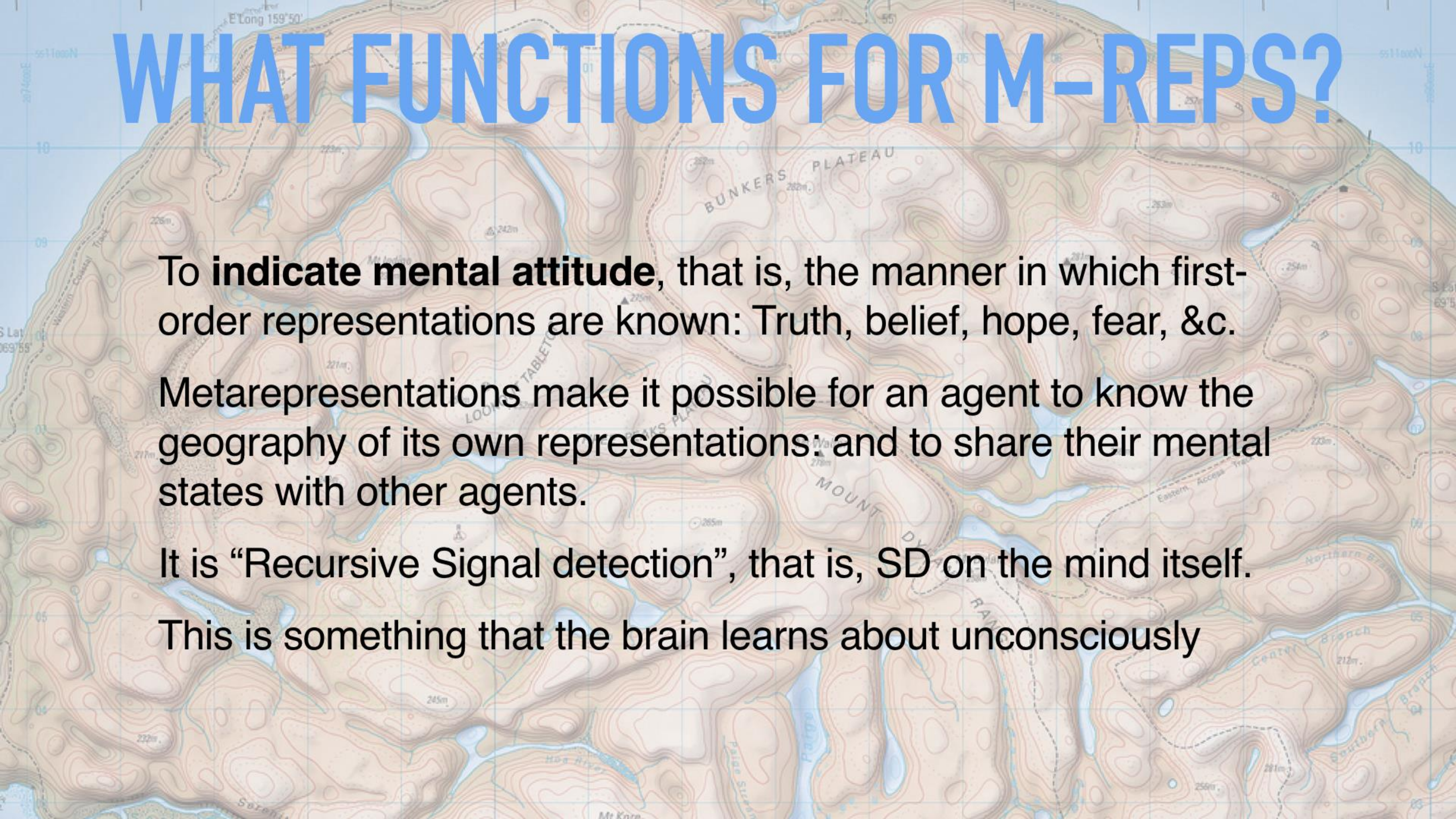
The knowledge is forever embedded in the causal chain implemented by the network. All the network can do is project this knowledge onto action. It is knowledge “in the network” vs. knowledge “for the network” (Clark & Karmiloff-Smith)



170 D. SERVAN-SCHREIBER, A. CLEEREMANS AND J.L. MCCLELLAND

Figure 8A. Hierarchical Cluster Analysis of the H.U. activation patterns after 200,000 presentations from strings generated at random according to the Reber grammar (Three hidden units).





WHAT FUNCTIONS FOR M-REPS?

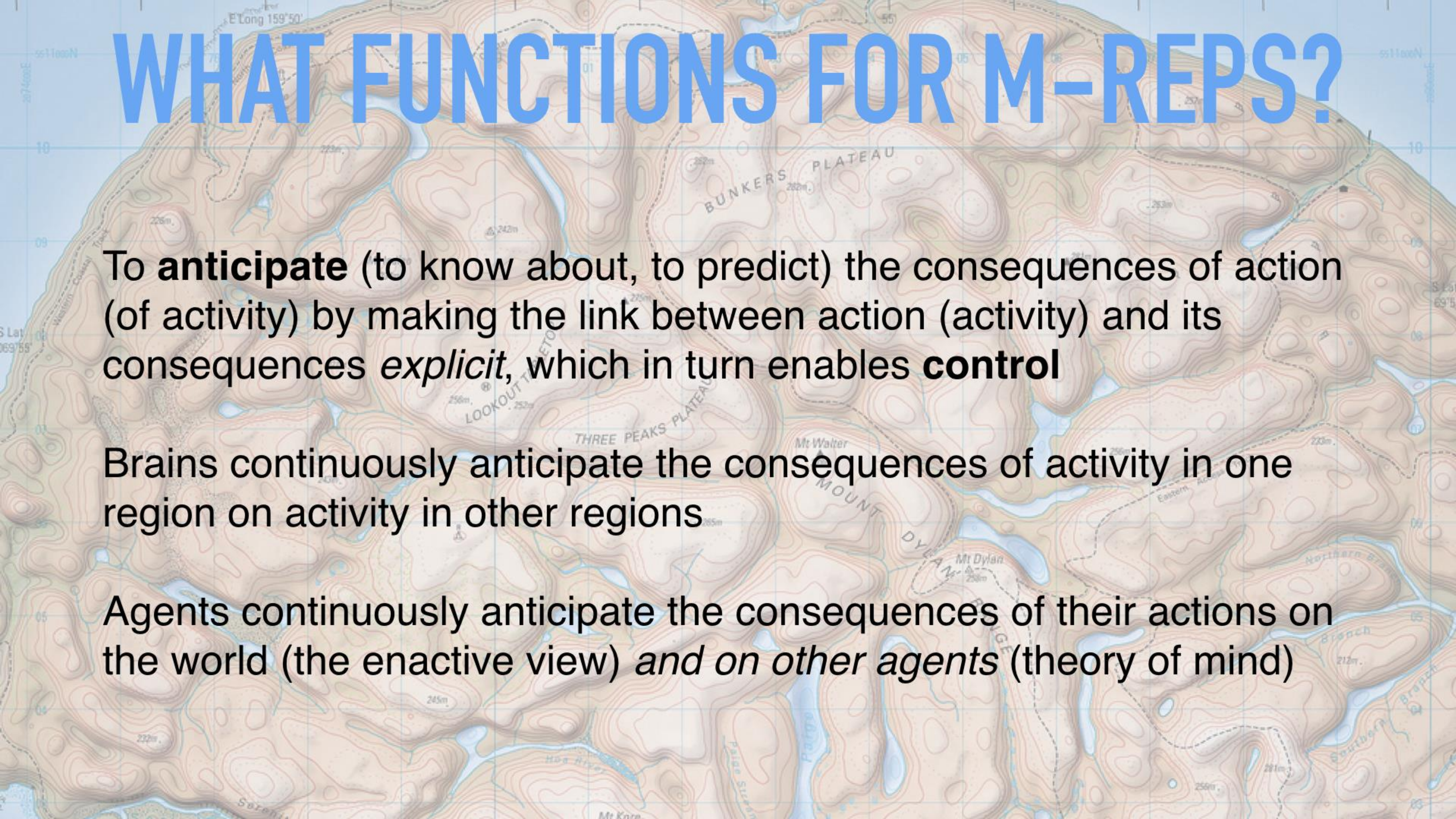
To **indicate mental attitude**, that is, the manner in which first-order representations are known: Truth, belief, hope, fear, &c.

Metarepresentations make it possible for an agent to know the geography of its own representations: and to share their mental states with other agents.

It is “Recursive Signal detection”, that is, SD on the mind itself.

This is something that the brain learns about unconsciously

WHAT FUNCTIONS FOR M-REPS?

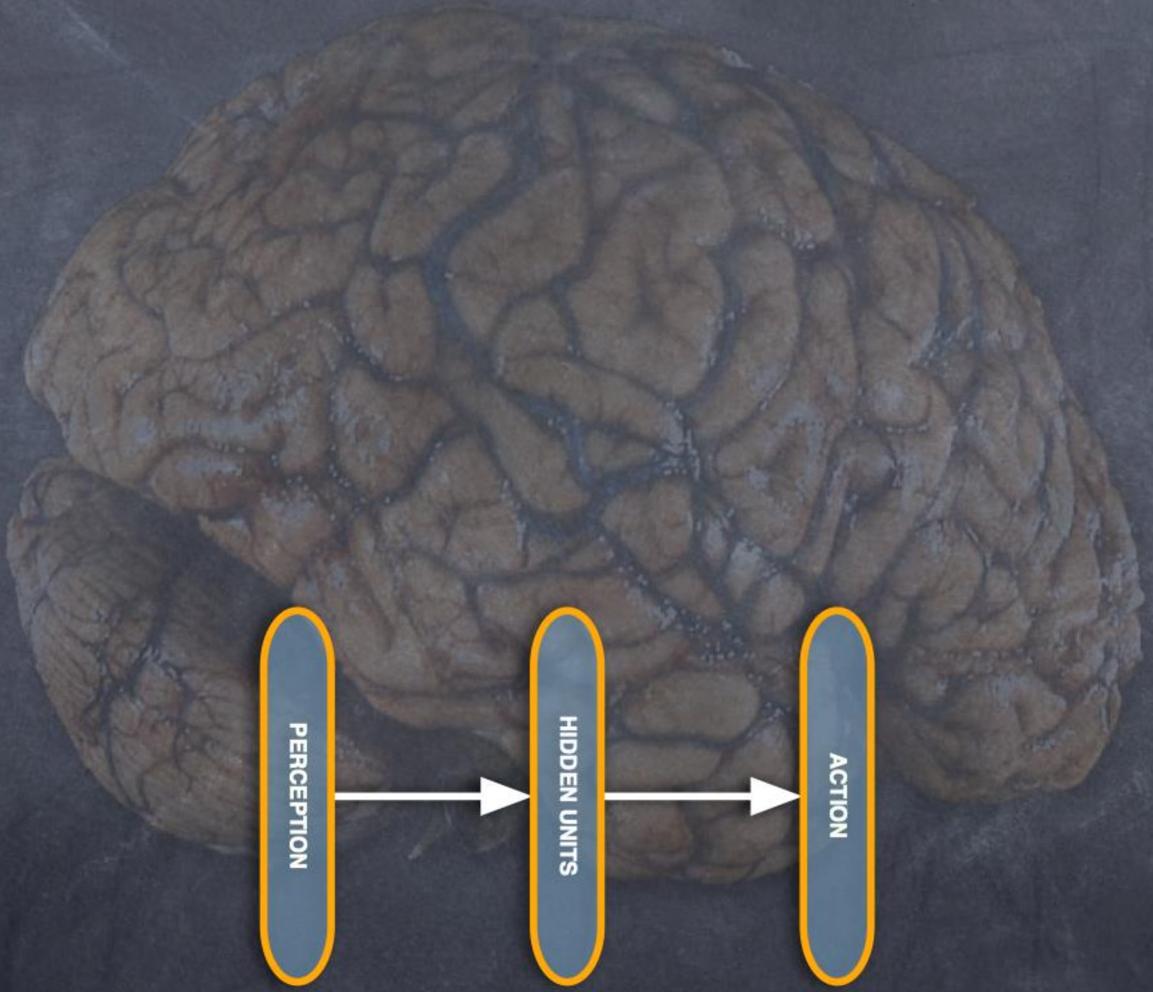


To **anticipate** (to know about, to predict) the consequences of action (of activity) by making the link between action (activity) and its consequences *explicit*, which in turn enables **control**

Brains continuously anticipate the consequences of activity in one region on activity in other regions

Agents continuously anticipate the consequences of their actions on the world (the enactive view) *and on other agents* (theory of mind)

SIGNAL DETECTION ON THE MIND

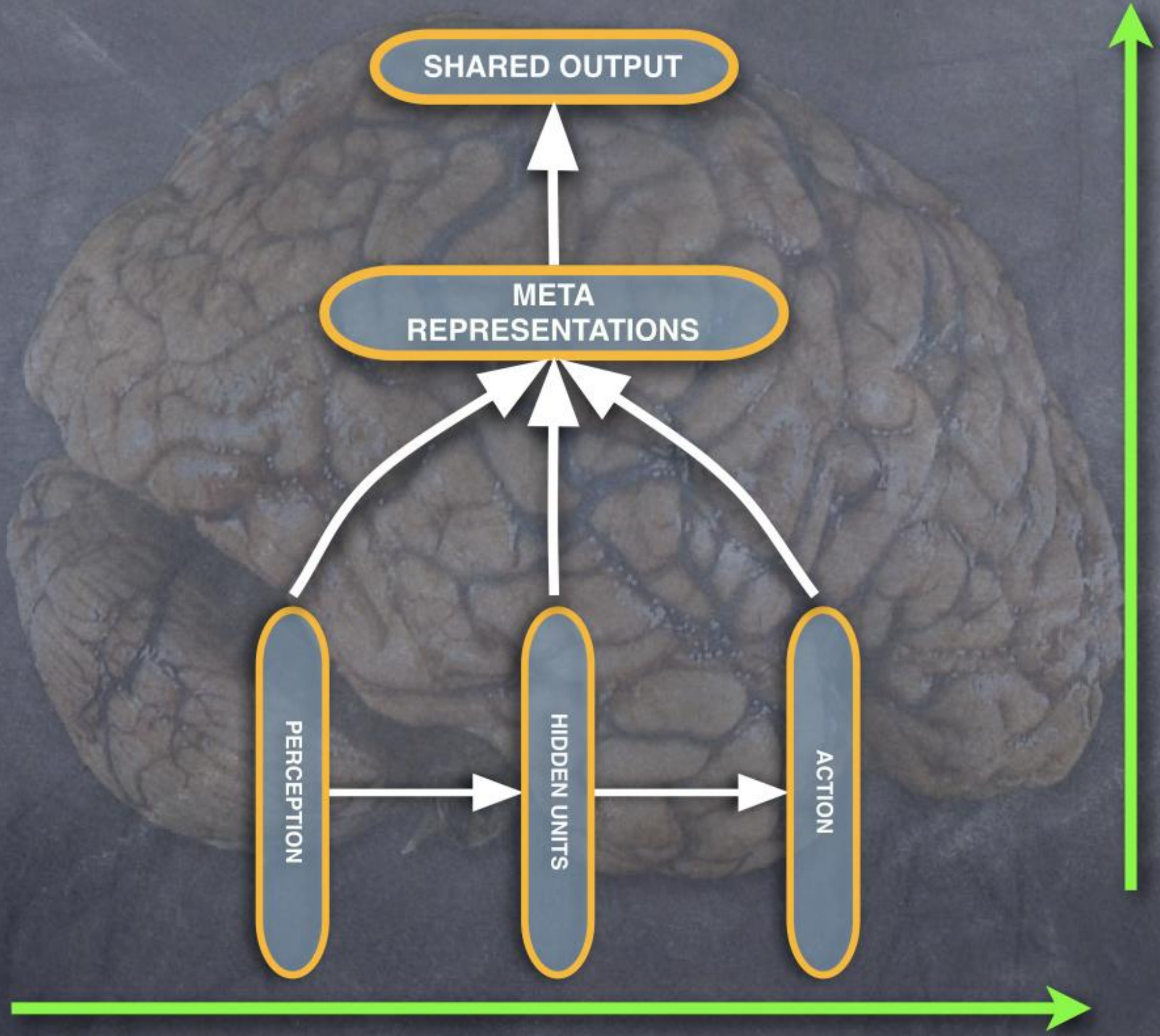


PERCEPTION

HIDDEN UNITS

ACTION

SIGNAL DETECTION ON THE MIND



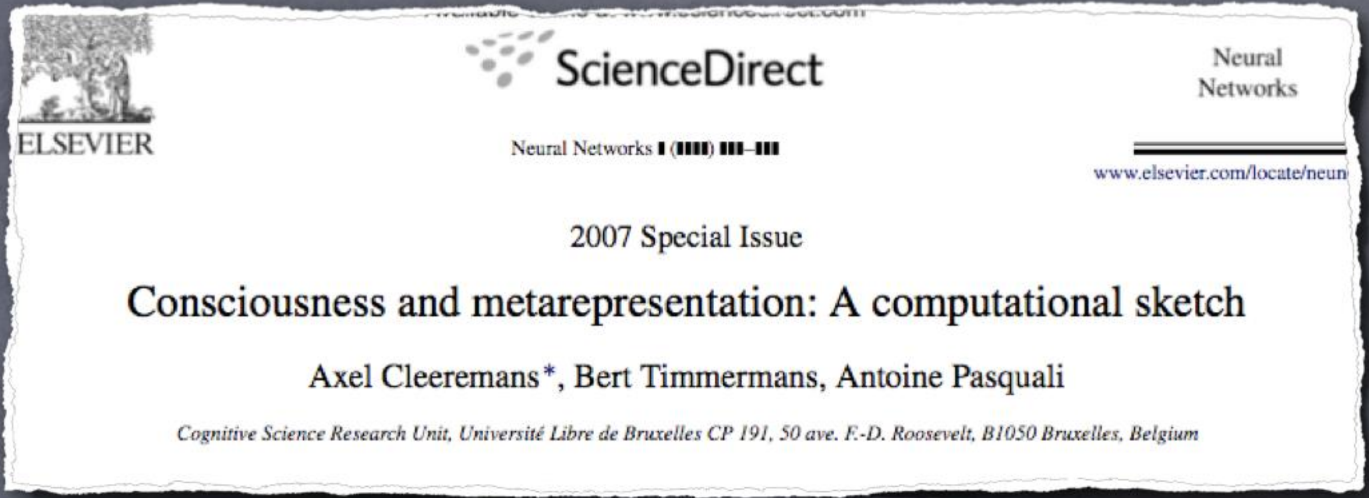
The brain learning about itself: Signal detection on your own representations

Type II decisions

Subjective measures

The brain learning about the world — Type I decisions, objective measures

WAGERING AS A MEASURE OF C



Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Neural Networks

Neural Networks ▮ (▮▮▮▮) ▮▮▮–▮▮▮

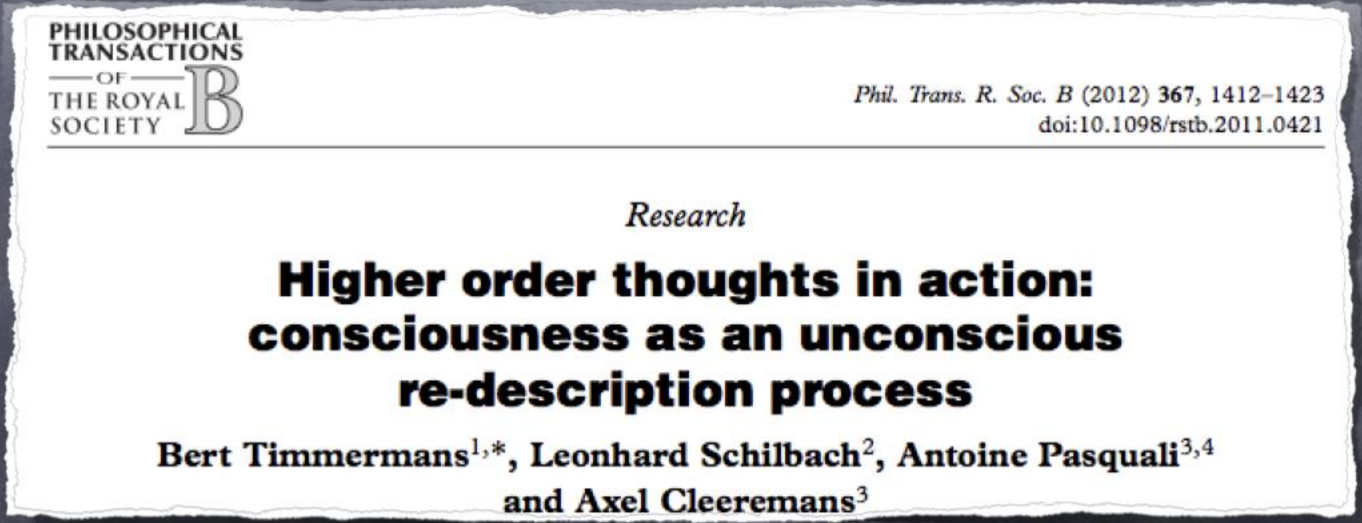
www.elsevier.com/locate/neun

2007 Special Issue

Consciousness and metarepresentation: A computational sketch

Axel Cleeremans*, Bert Timmermans, Antoine Pasquali

Cognitive Science Research Unit, Université Libre de Bruxelles CP 191, 50 ave. F.-D. Roosevelt, B1050 Bruxelles, Belgium



PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B

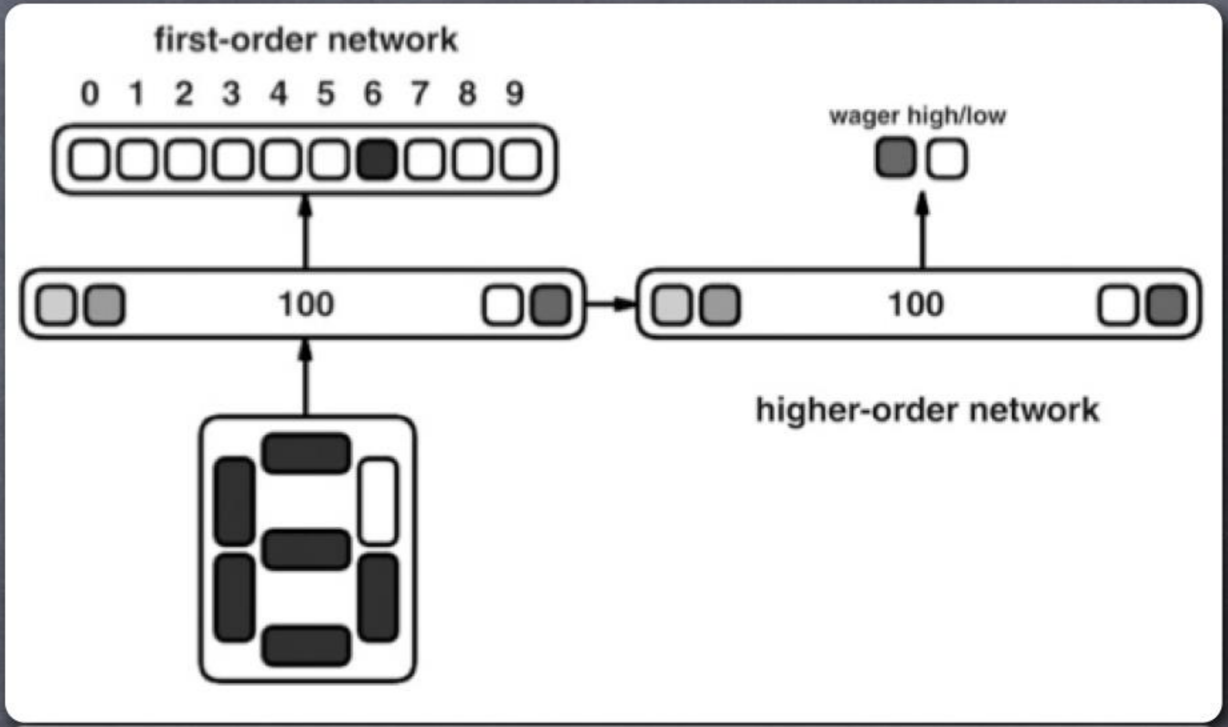
Phil. Trans. R. Soc. B (2012) **367**, 1412–1423
doi:10.1098/rstb.2011.0421

Research

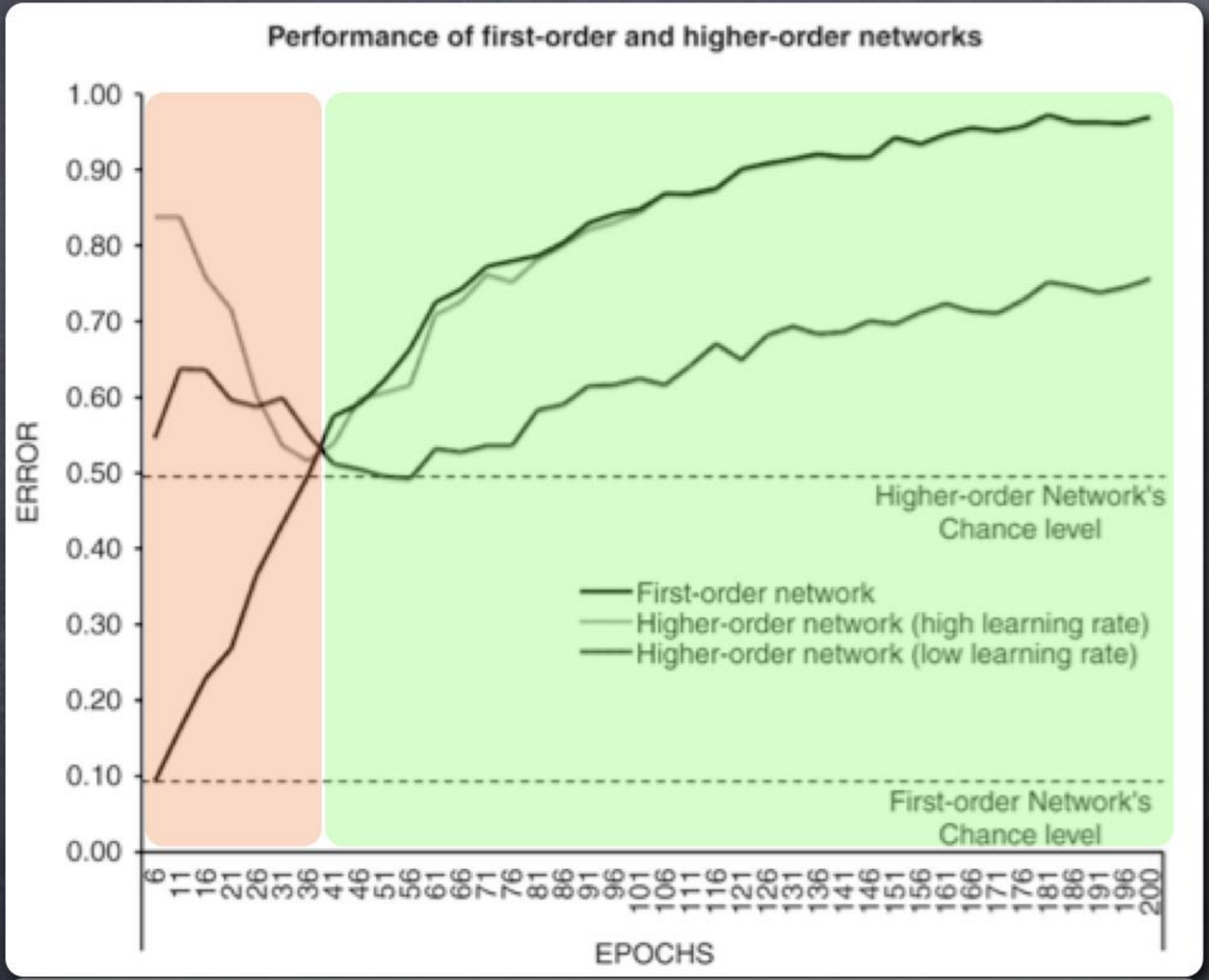
Higher order thoughts in action: consciousness as an unconscious re-description process

Bert Timmermans[1,*], Leonhard Schilbach[2], Antoine Pasquali[3,4] and Axel Cleeremans[3]

WAGERING IN THE DIGITS TASK



WAGERING IN THE DIGITS TASK

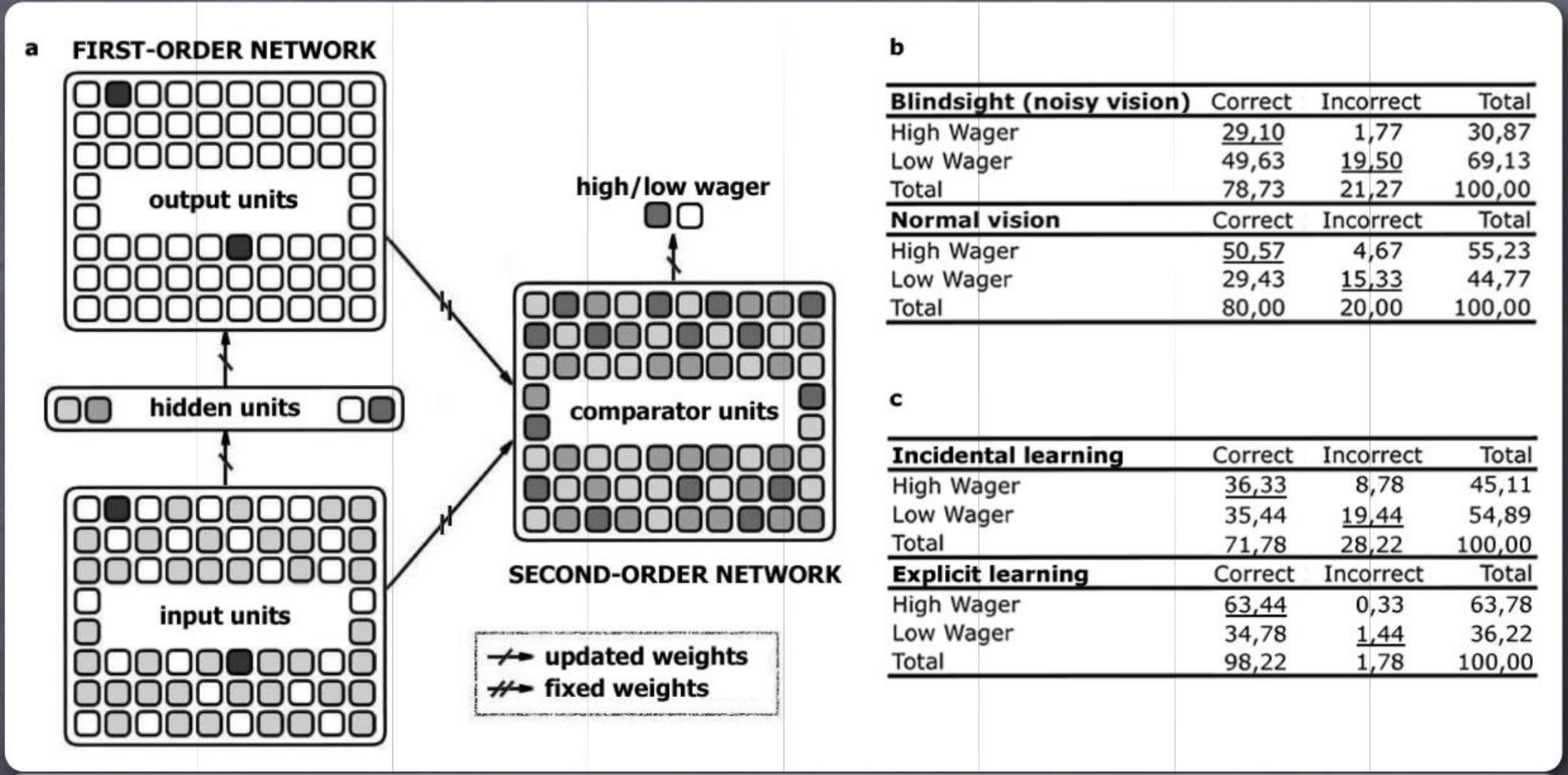


Different training conditions result in different patterns of relationship between the performance of the first-order network and that of the second-order (wagering) network

Early in training, the first-order network is performing well above chance, yet the second-order network's betting performance decreases and eventually gets close to chance level. This, by subjective measures, indicates unconscious processing.

Later, the performance of the two networks correlate, suggesting conscious knowledge.

BLINDSIGHT & IMPLICIT LEARNING



b

Blindsight (noisy vision)	Correct	Incorrect	Total
High Wager	29,10	1,77	30,87
Low Wager	49,63	19,50	69,13
Total	78,73	21,27	100,00
Normal vision	Correct	Incorrect	Total
High Wager	50,57	4,67	55,23
Low Wager	29,43	15,33	44,77
Total	80,00	20,00	100,00

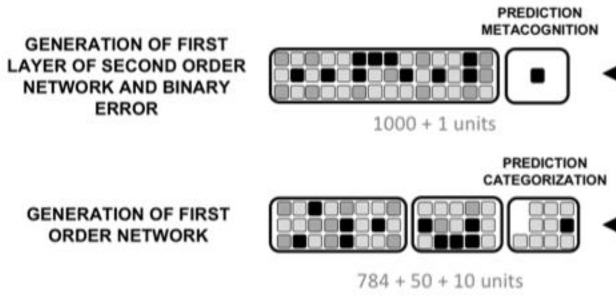
c

Incidental learning	Correct	Incorrect	Total
High Wager	36,33	8,78	45,11
Low Wager	35,44	19,44	54,89
Total	71,78	28,22	100,00
Explicit learning	Correct	Incorrect	Total
High Wager	63,44	0,33	63,78
Low Wager	34,78	1,44	36,22
Total	98,22	1,78	100,00

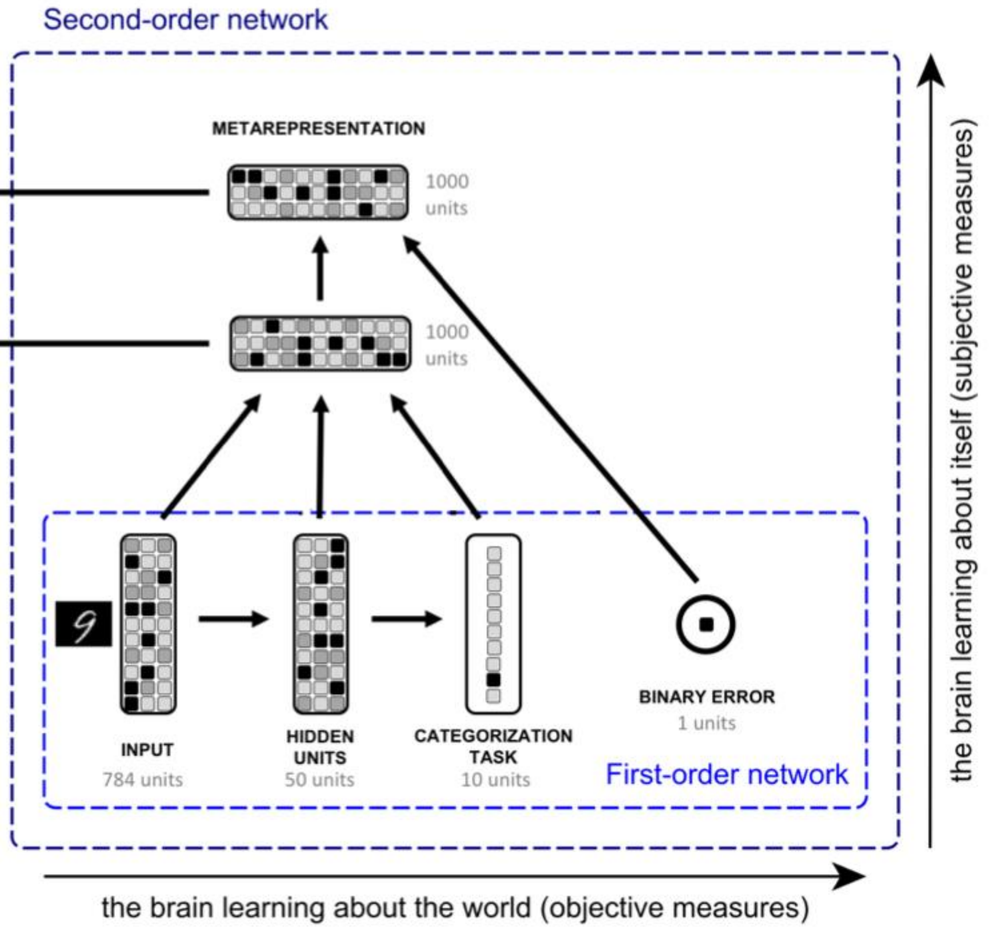
RBM MODEL

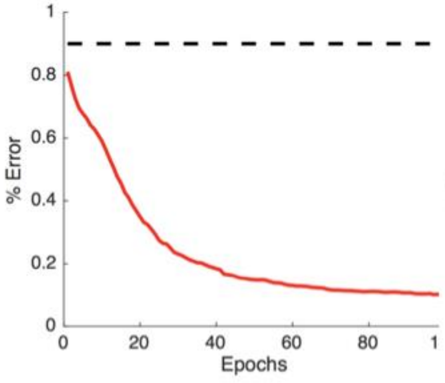
Arnaud Beauny

Type I task: categorize 1 digit among 10
Type II task: Is the first order network right or wrong ?

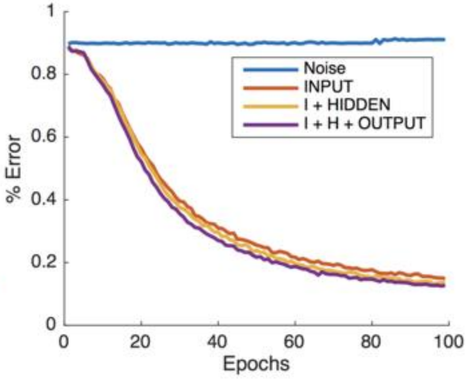


GENERATION OF FIRST LAYER OF SECOND ORDER NETWORK AND BINARY ERROR
PREDICTION METACOGNITION
1000 + 1 units
GENERATION OF FIRST ORDER NETWORK
PREDICTION CATEGORIZATION
784 + 50 + 10 units

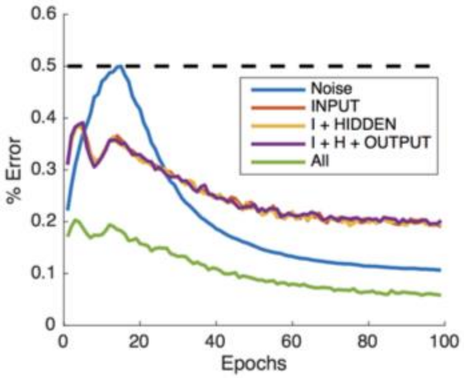




Categorization task (BackProp)

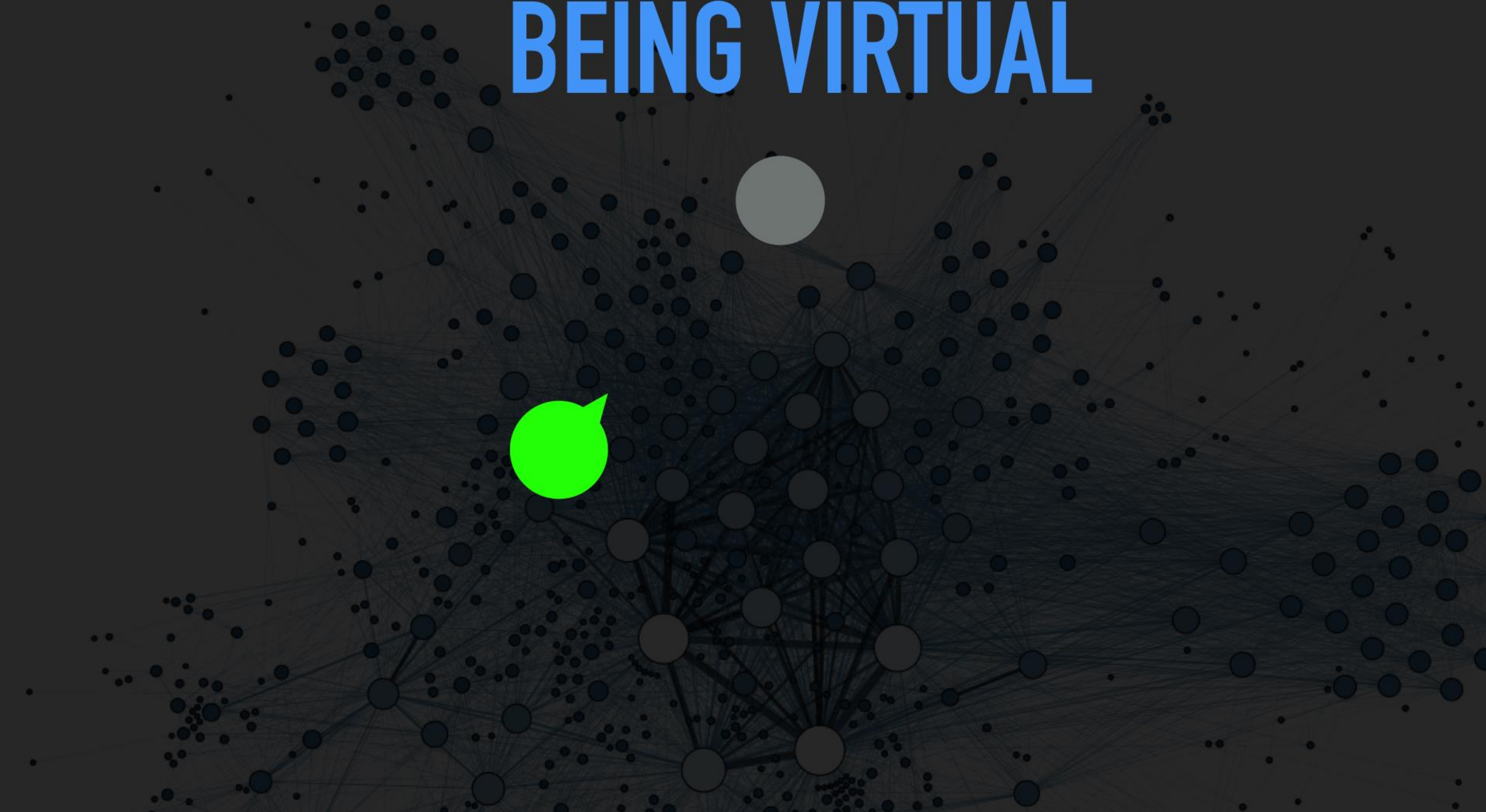


Prediction categorization (RBM)

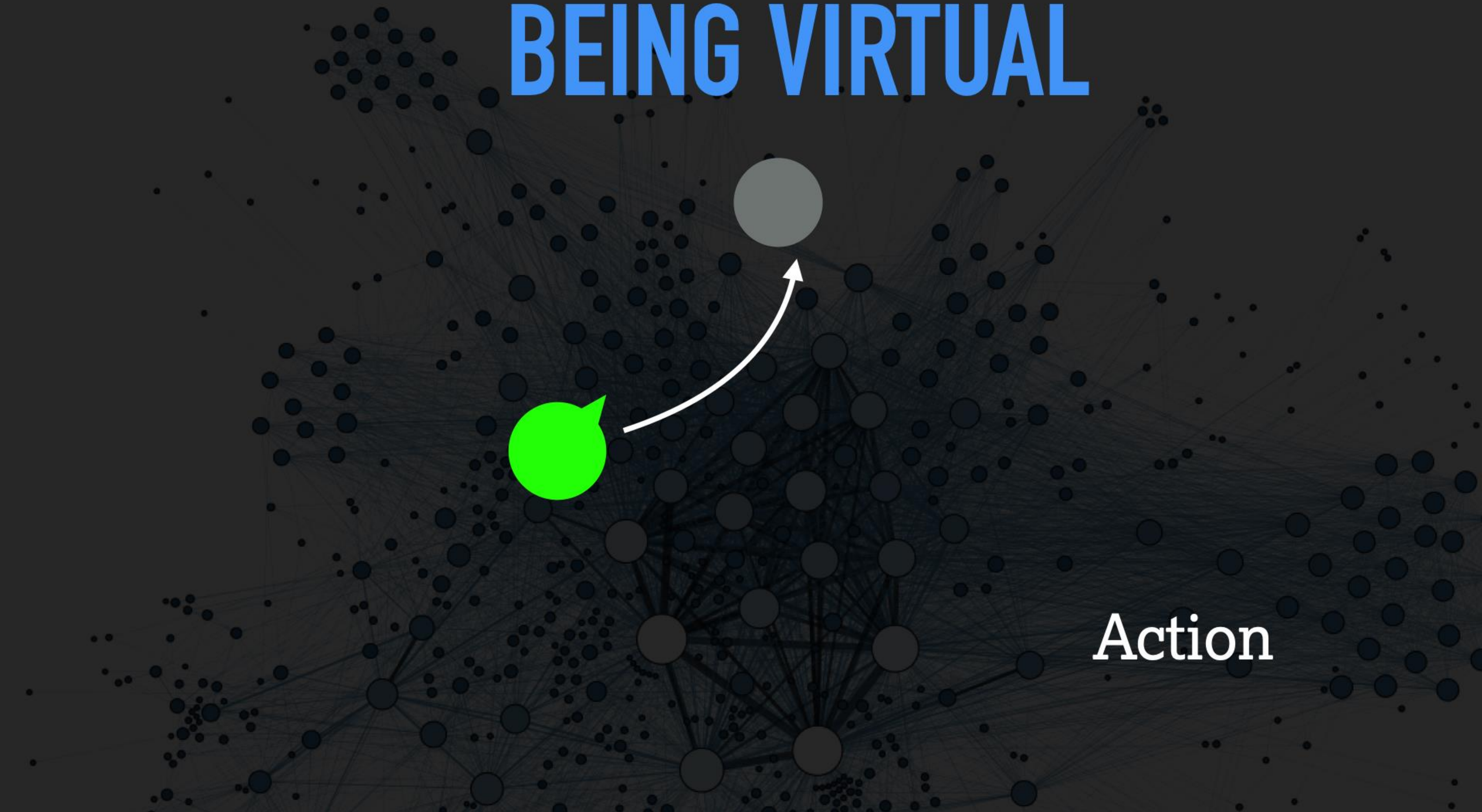


Prediction metacognition (RBM)

BEING VIRTUAL

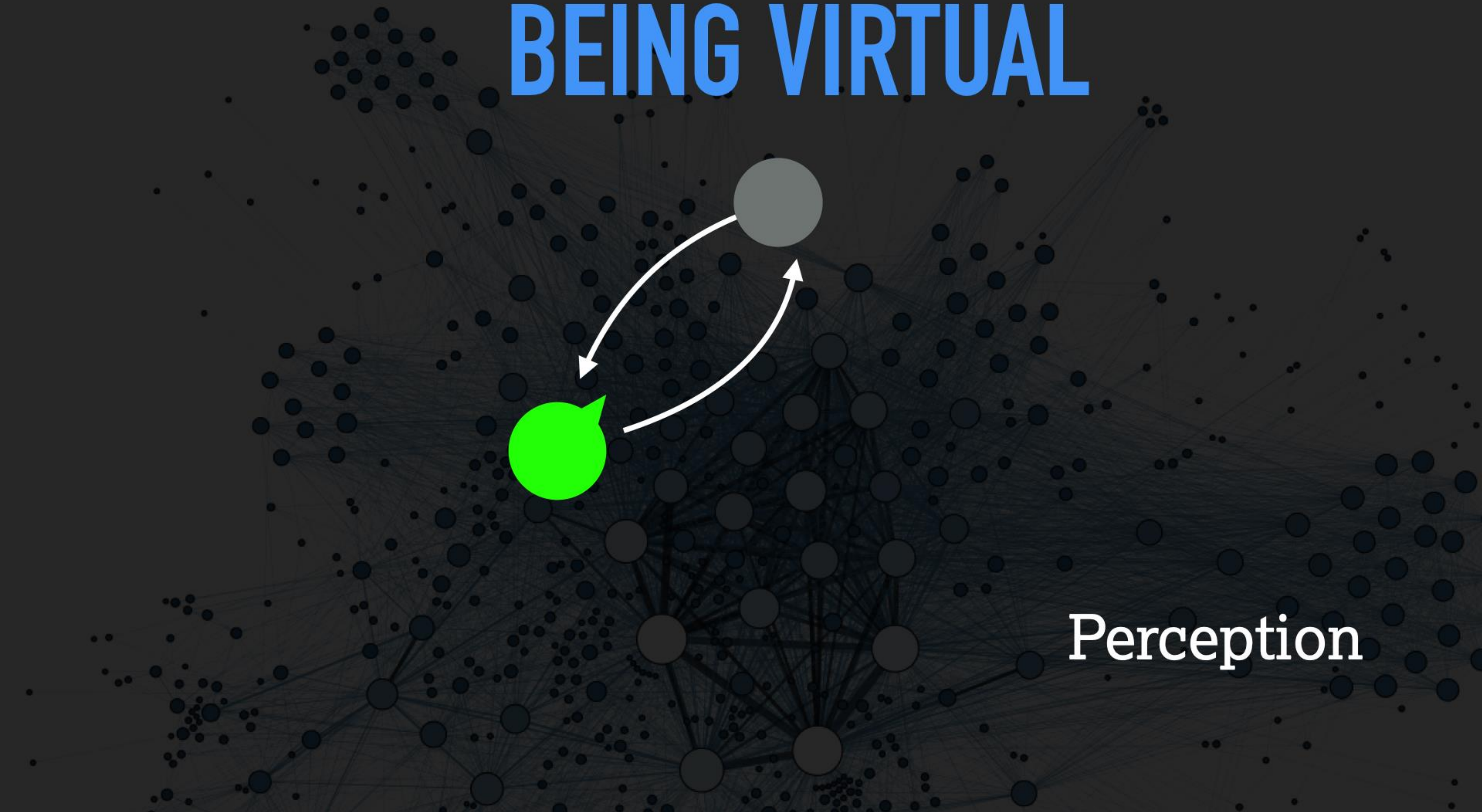


BEING VIRTUAL



Action

BEING VIRTUAL

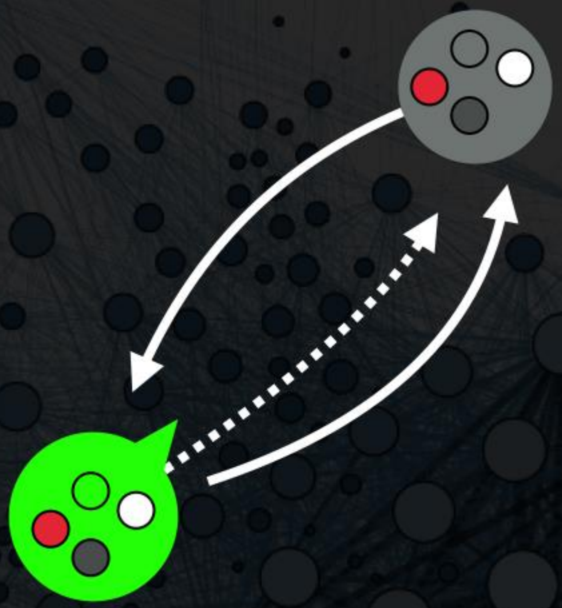


Perception

BEING VIRTUAL

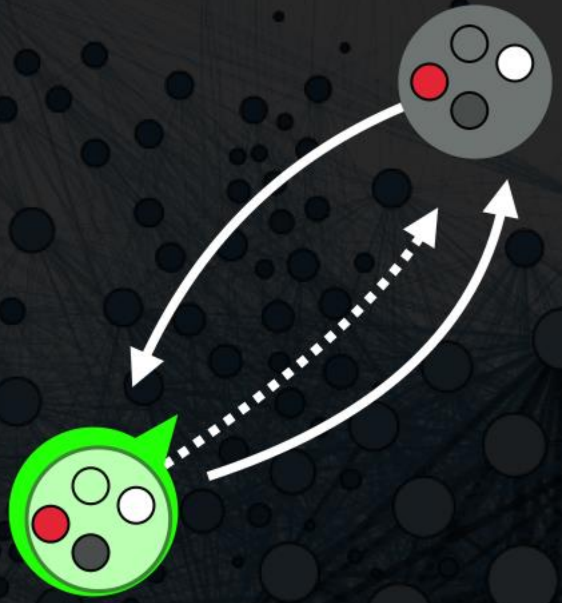
Prediction

BEING VIRTUAL



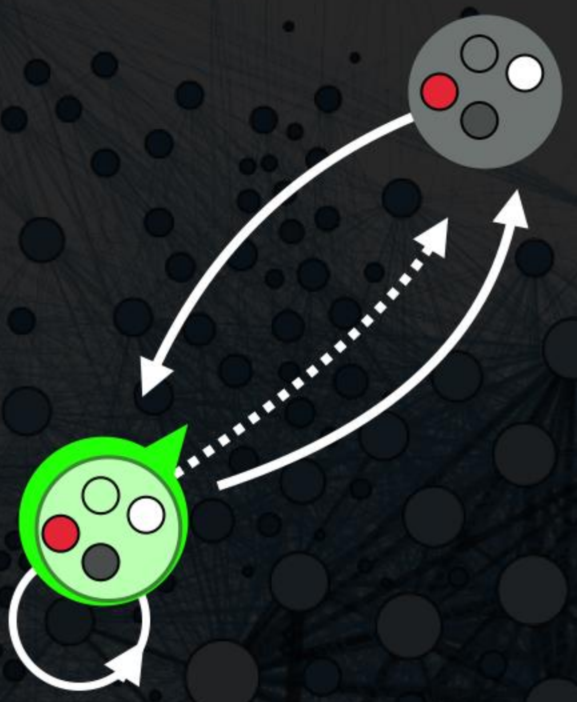
“Inference”

BEING VIRTUAL



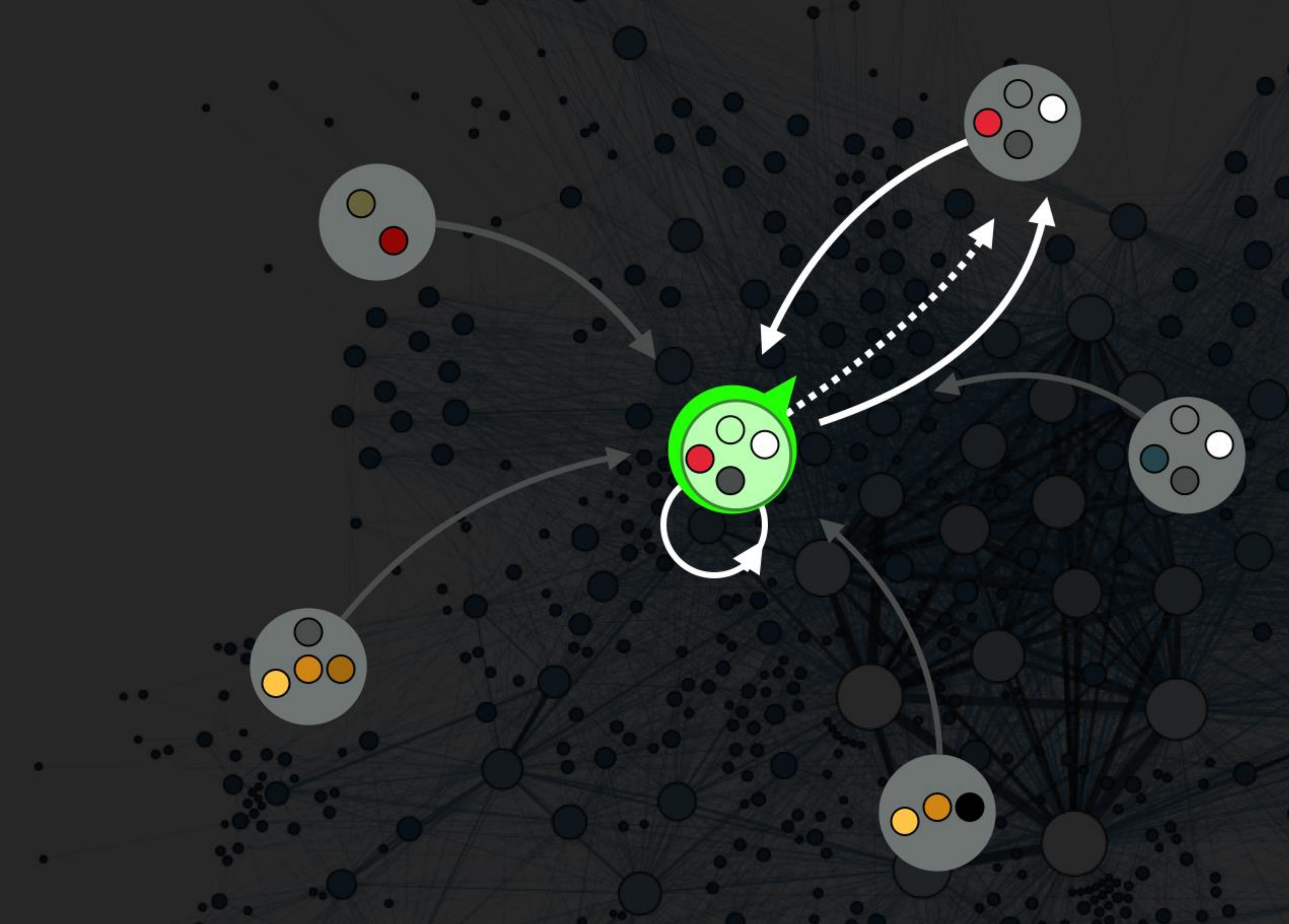
Now the agent has
built a model of
what
it is to be an agent

BEING VIRTUAL



… which it can use as a model of itself

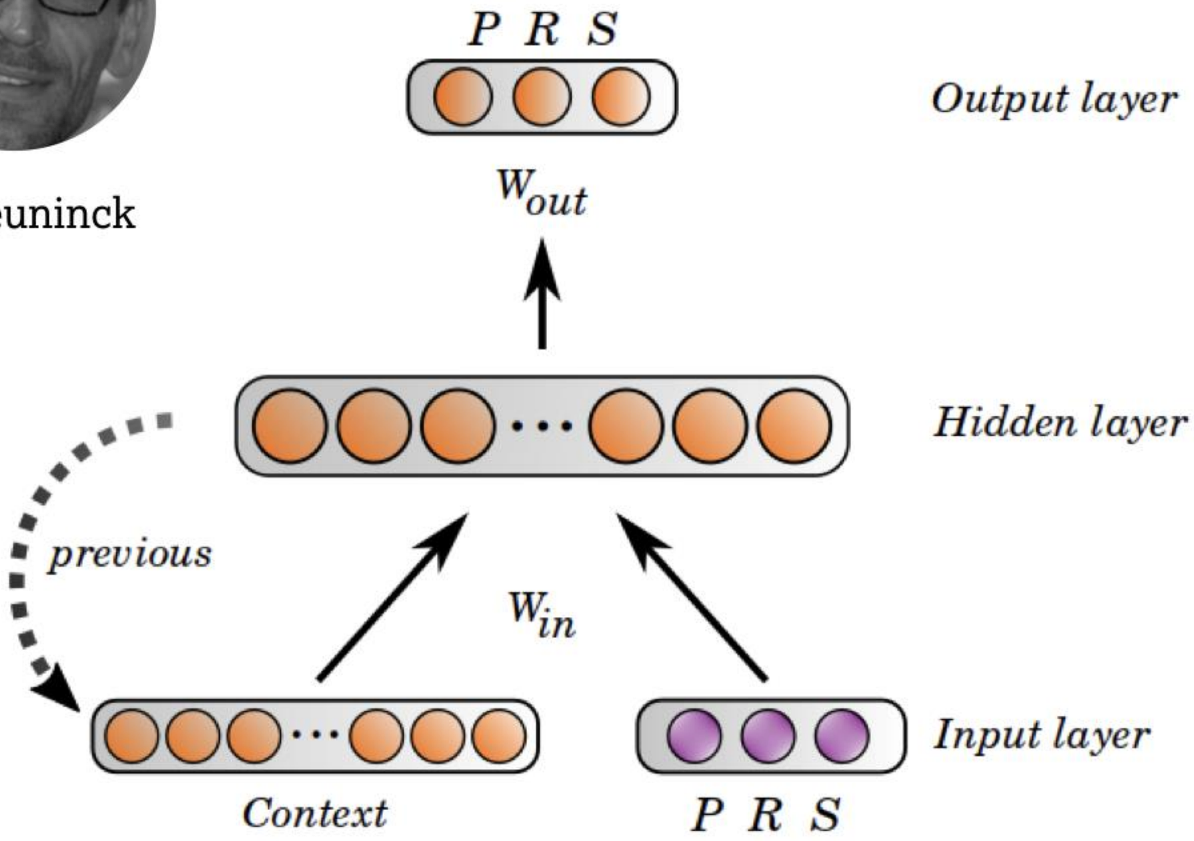
BEING VIRTUAL



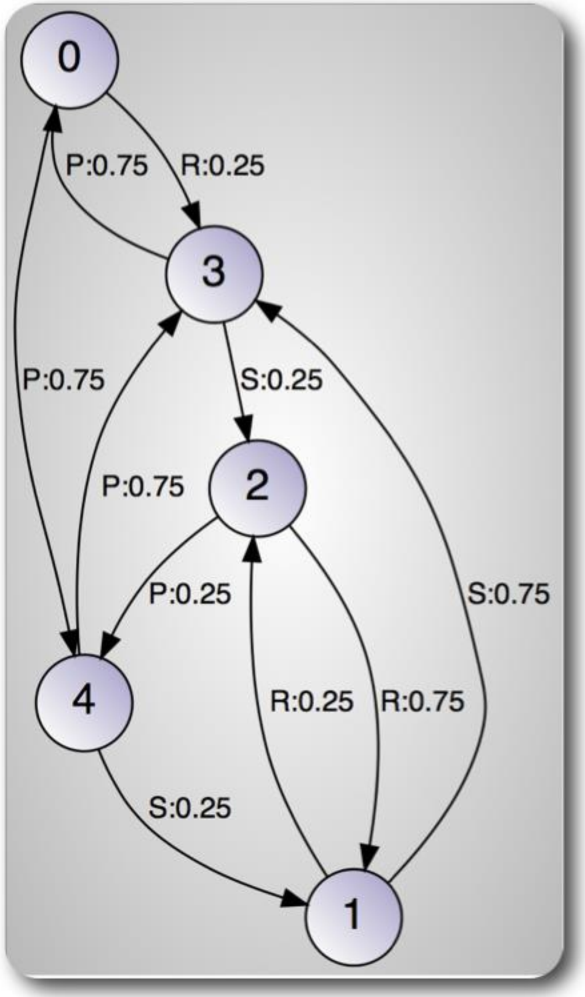
This repeats many times with other agents

WHAT IT IS LIKE TO BE A FINITE-STATE MACHINE?

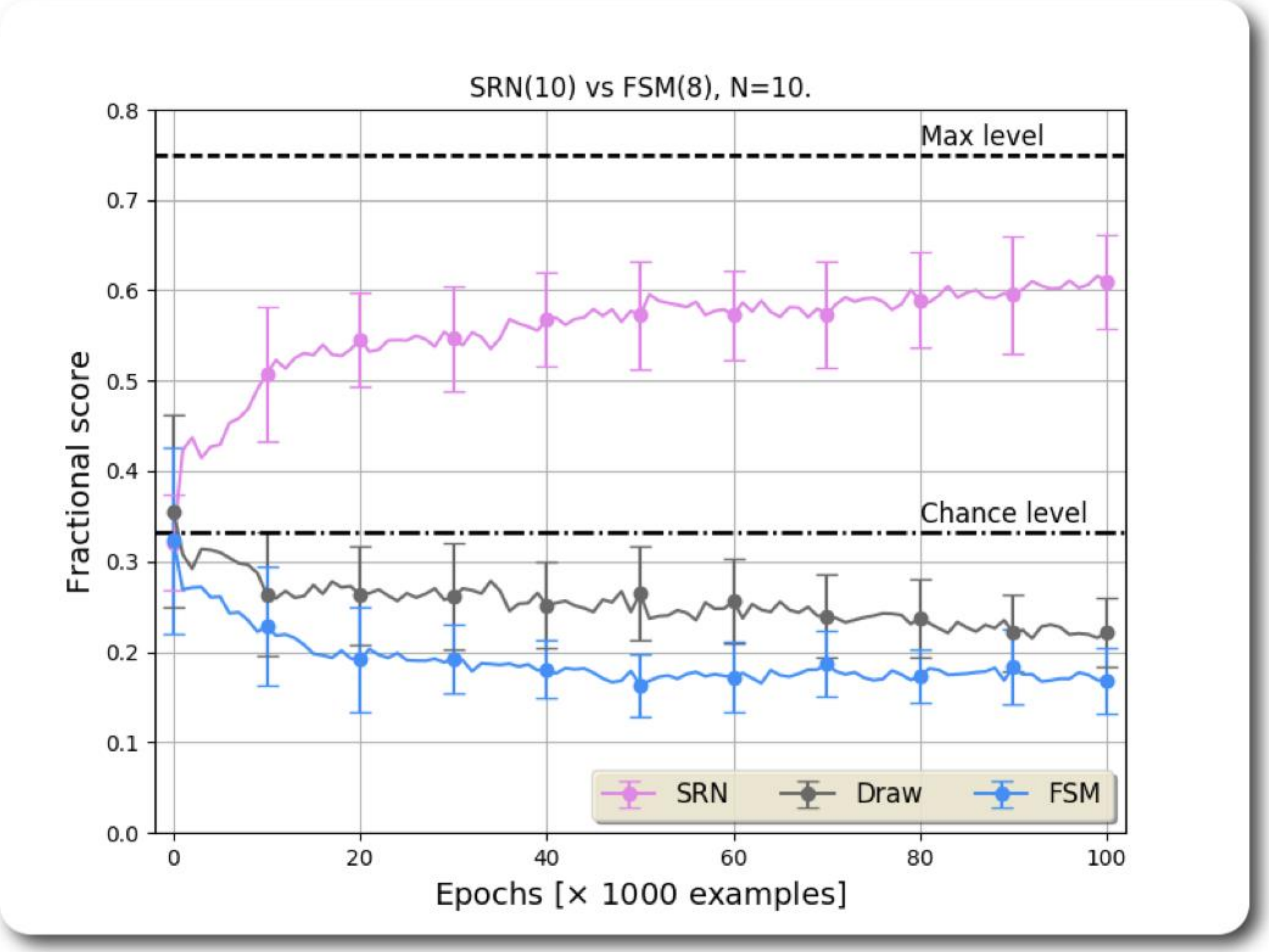
Lars Keuninck



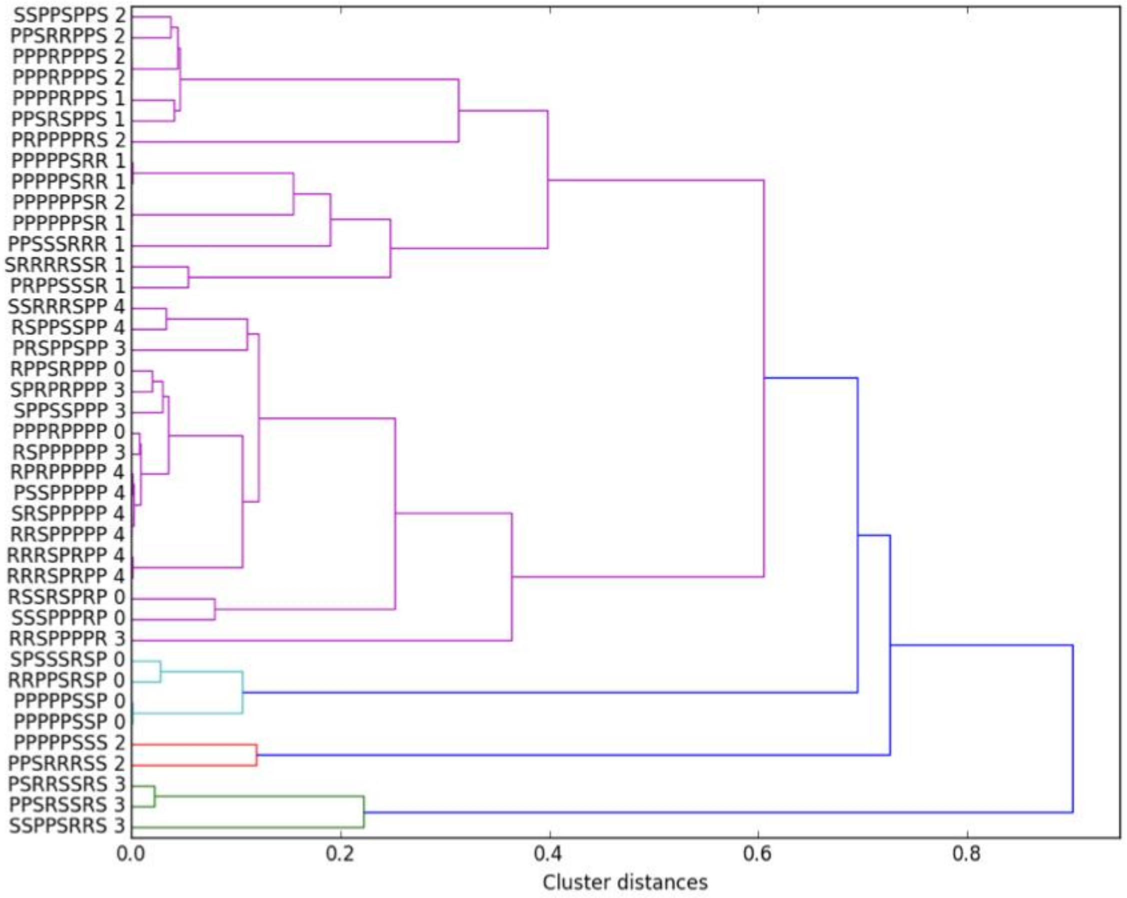
Cleeremans et al. (1989), Elman (1990)

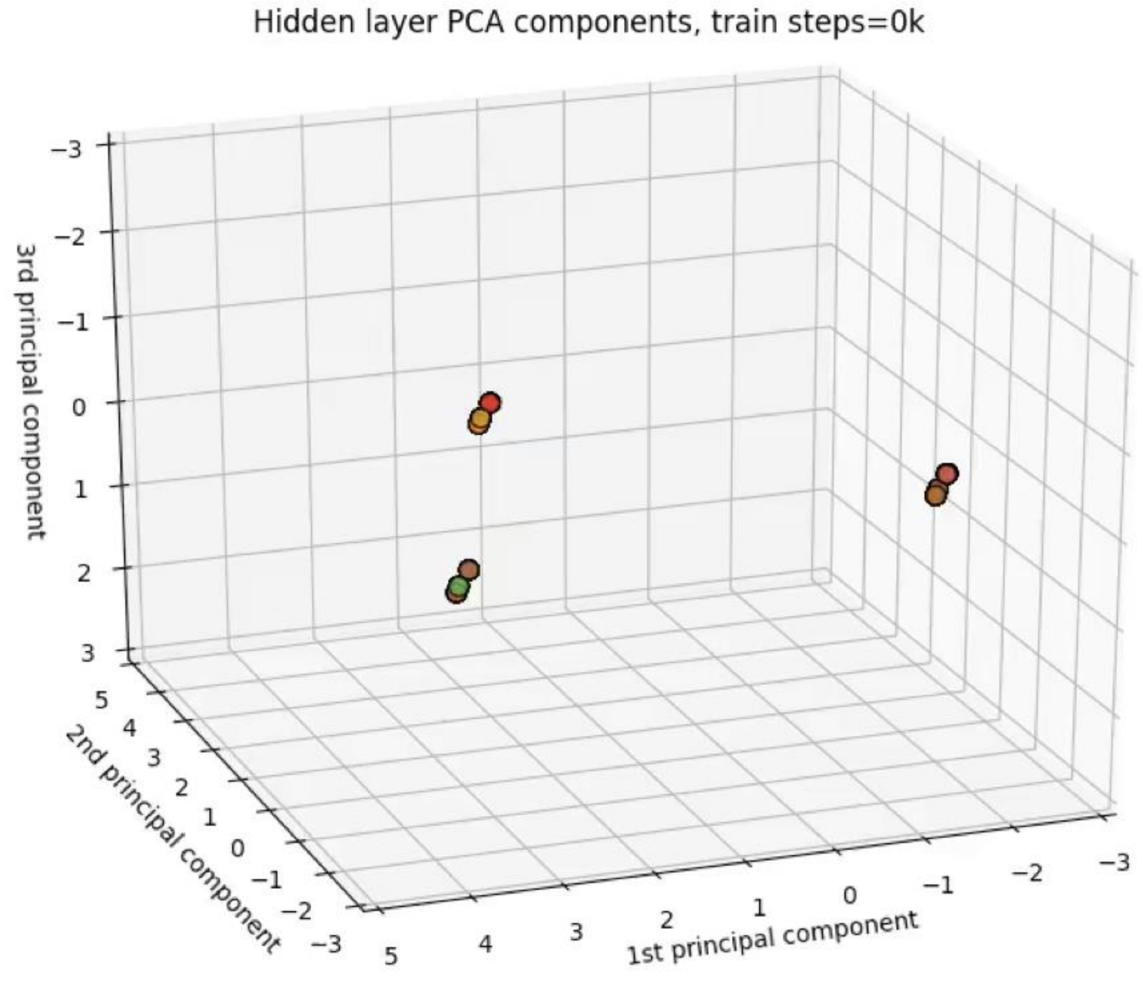


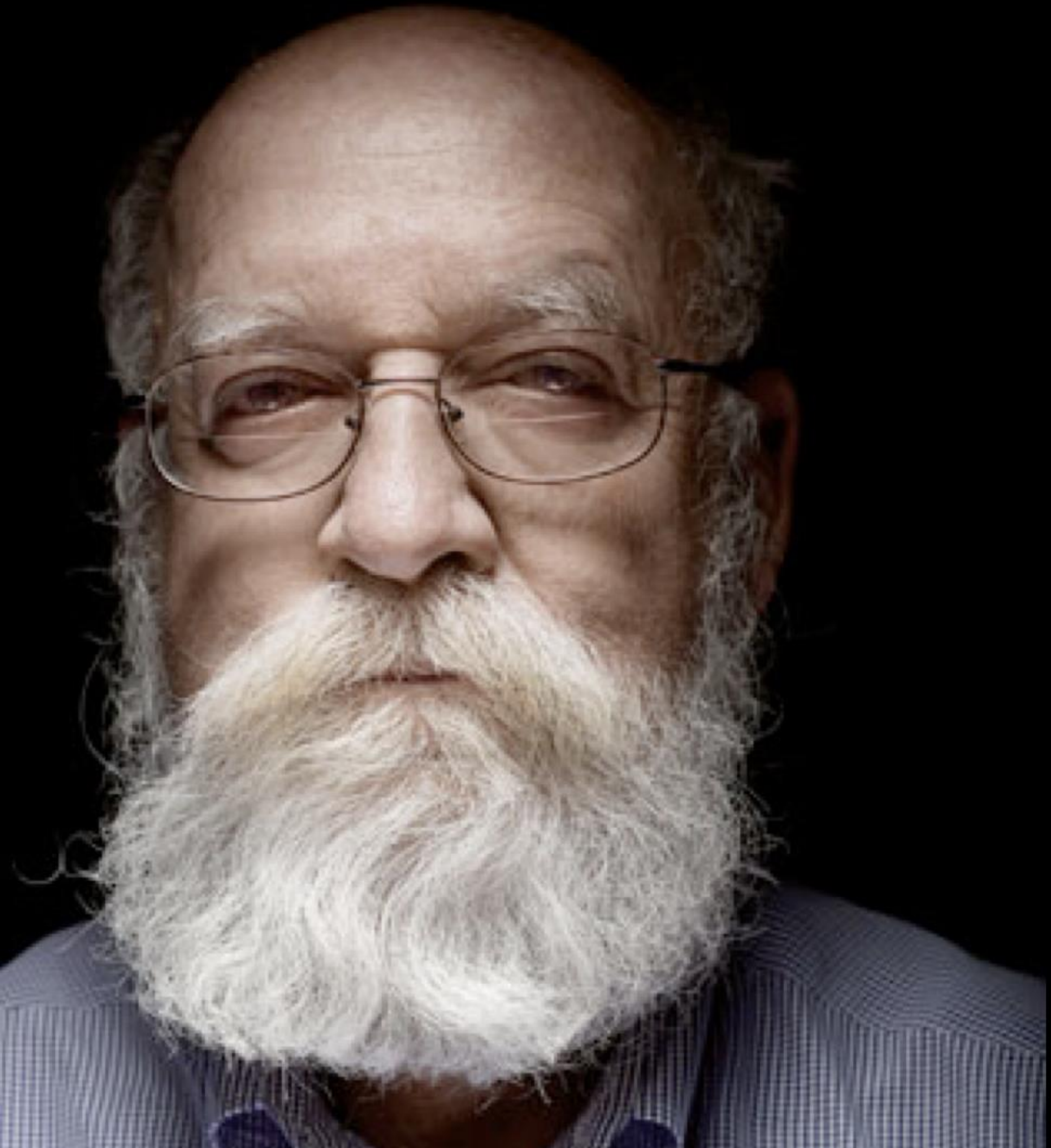
WHAT IT IS LIKE TO BE A FINITE-STATE MACHINE?



WHAT IT IS LIKE TO BE A FINITE-STATE MACHINE?







“How do we go from doing things for reasons to having reasons for doing things?”

WHO IS CONSCIOUS?

Three criteria?

- ★ Massive information-processing resources that are sufficiently powerful to simulate certain aspects of their own physical basis and inner workings;
- ★ A continuously learning system that attempts to predict future states;
- ★ Immersion in a sufficiently rich social environment from which models of yourself can be built.

CONCLUSIONS

- Consciousness is more than either “sensitivity” or intelligence
- Chalmers’ “hard problem” remains intact
- Consciousness is the brain’s (unconscious, enactive, embodied) theory about itself
- There is no principled argument against the possibility of conscious machines
- Contemporary intelligent artefacts lack agenthood: They neither want anything nor care about anything
- Do we really want to build conscious, superhuman intelligent agents that are also immortal and infinitely replicable?

ULB

CENTER FOR RESEARCH IN COGNITION & NEUROSCIENCES /CO3

+ ESTI SAN ANTON

crcn.ulb.ac.be/co3

fnrs LA LIBERTÉ DE CHERCHER

erc European Research Council Established by the European Commission

CIFAR CANADIAN INSTITUTE FOR ADVANCED RESEARCH

Belgian Science Policy Office belspo

FUNDAÇÃO Bial Institution of public utility

ULB NEUROSCIENCE INSTITUTE