

Computation, cognition, and control

Owen Holland



University of Essex

Department of
Computer Science

BICS

Brains

Kangaroos

Cognitive Systems

Autonomous agents

Simulation

Conclusions

BICS

Brains

Kangaroos

Cognitive Systems

Autonomous agents

Simulation

Conclusions

How to build a conscious machine

Why we shouldn't even try

BICS

BICS

Brain Inspired Cognitive Systems

BICS

Brain Inspired Cognitive Systems

What is a brain?

BICS

Brain Inspired Cognitive Systems

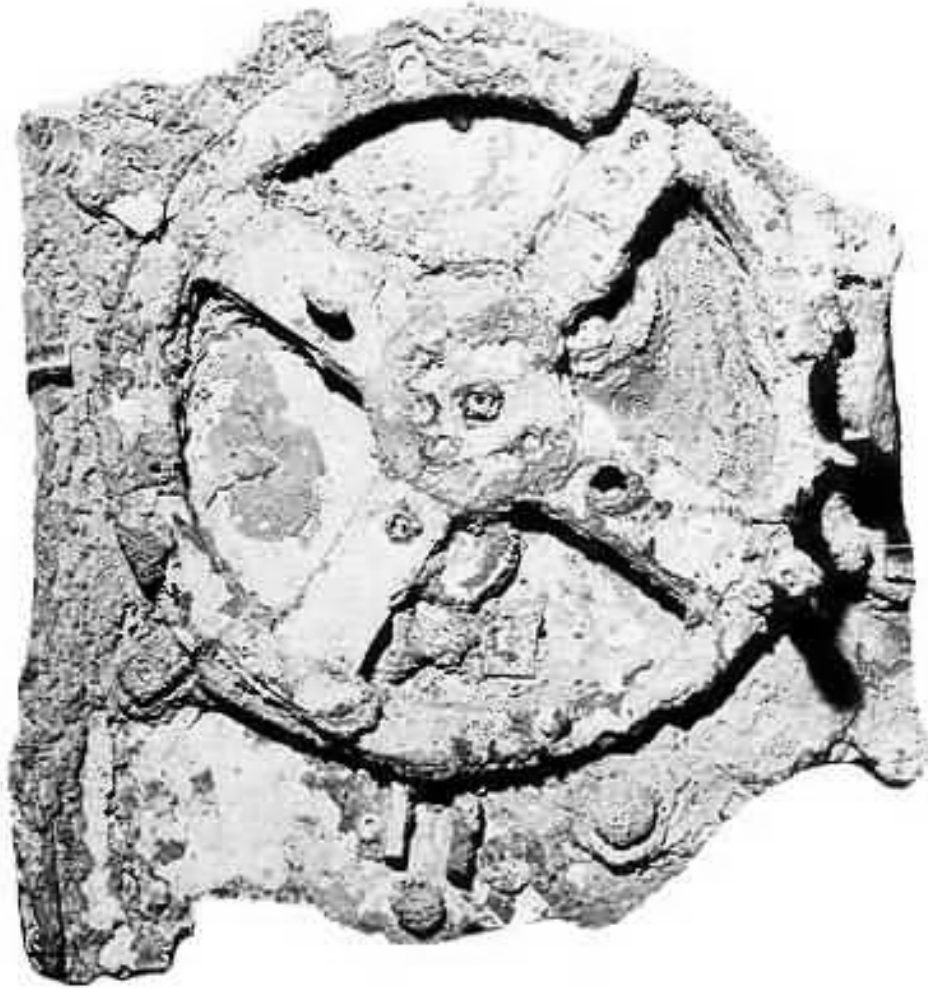
What is a brain?

What is a cognitive system?

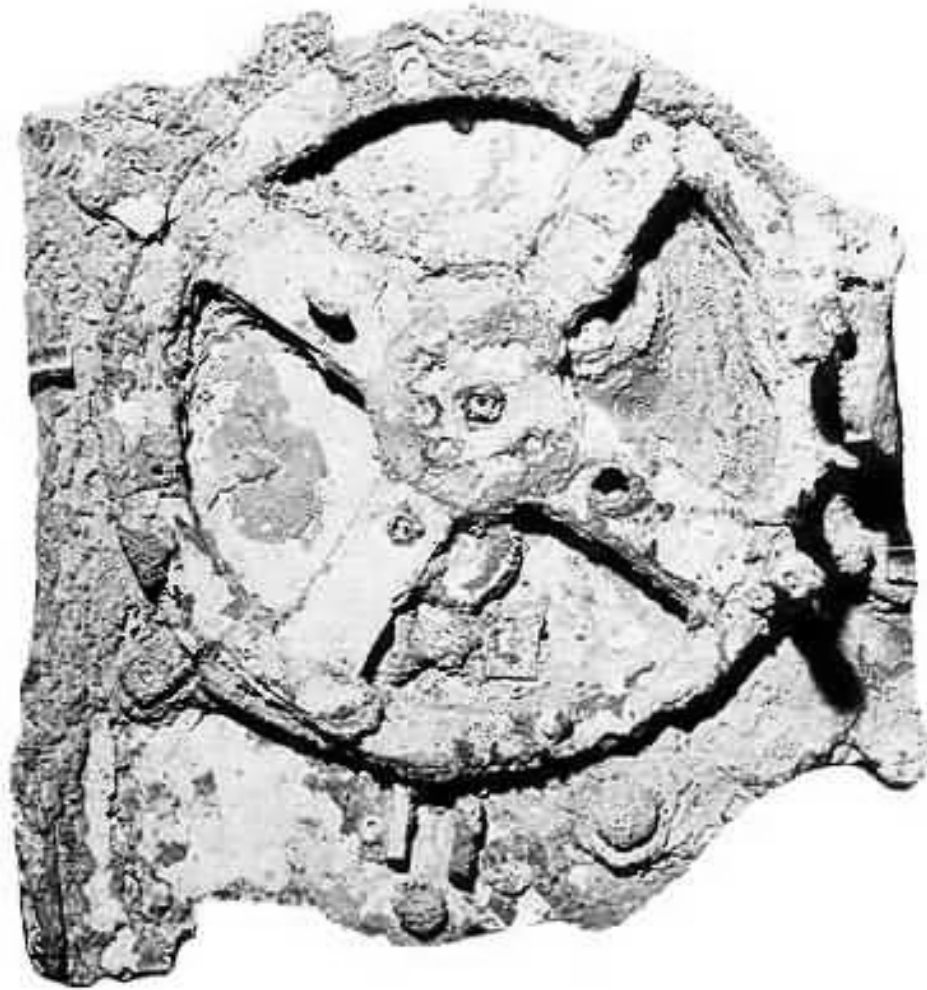
What is this?



What is it for?



How does it work?



How does it work?



If you knew what it was supposed to do, you could work out how it did it

What is this?



What is it for?



How does it work?

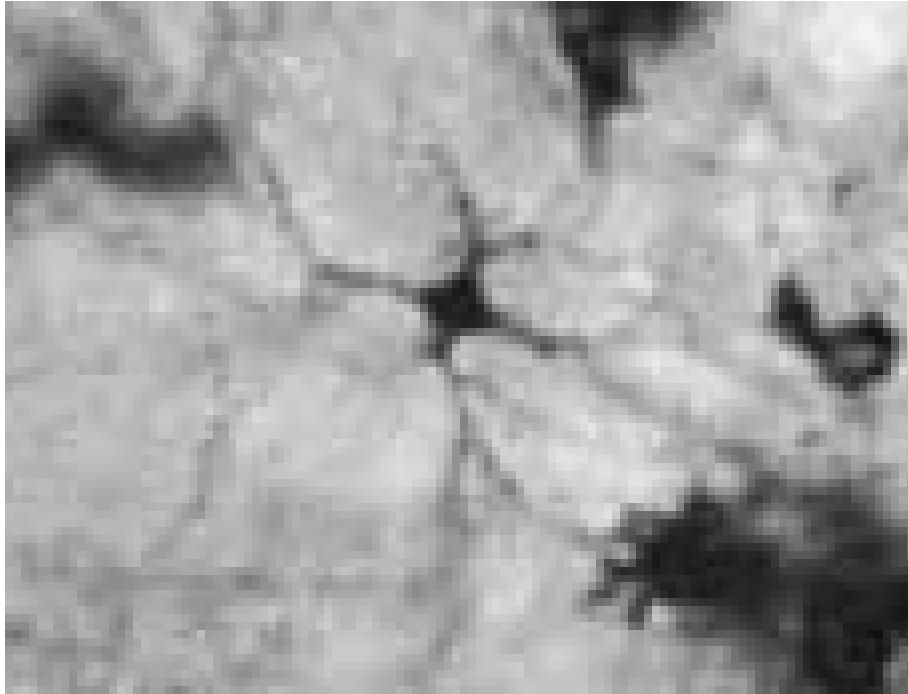


How does it work?

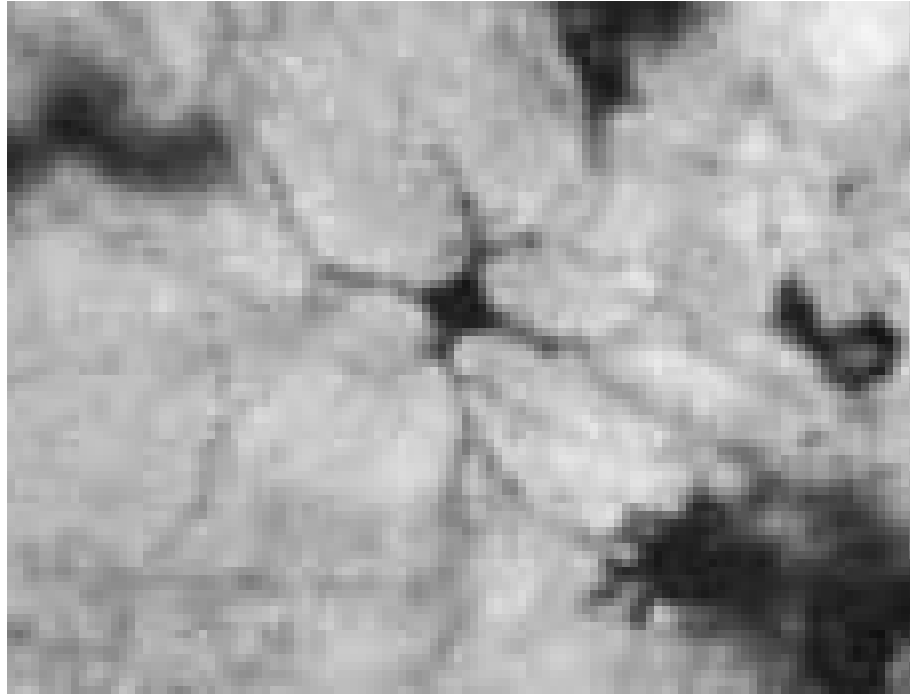


If you knew what it was supposed to do, you could work out how it did it...?

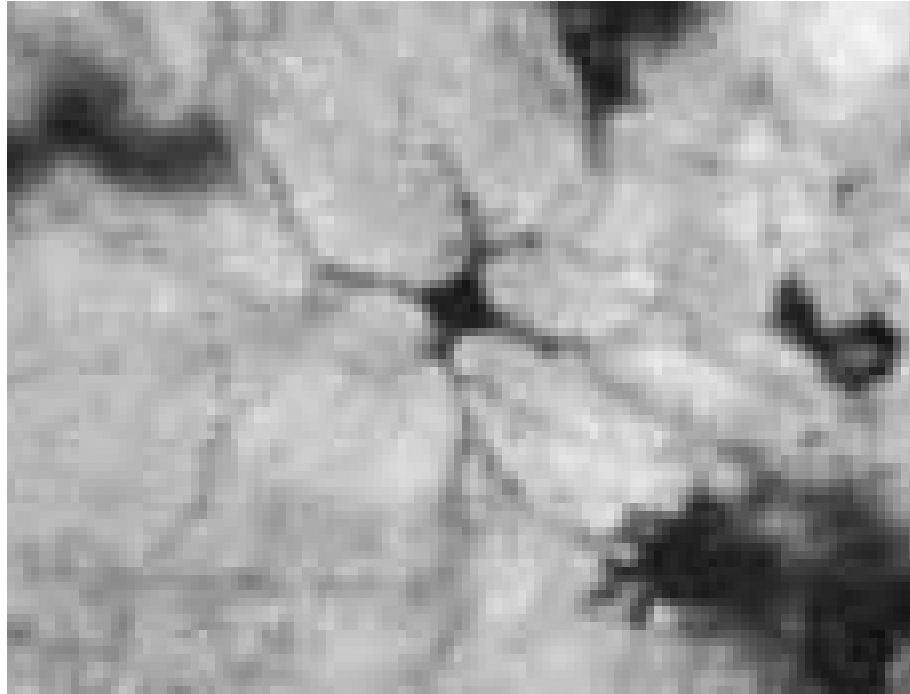
What is this?



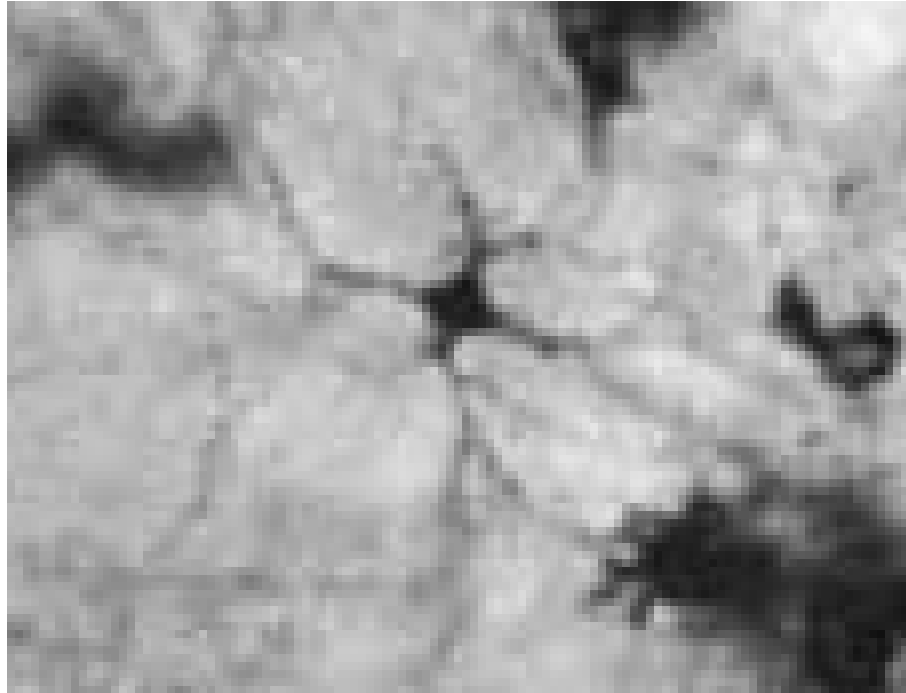
What is it for?



How does it work?



How does it work?



If you knew what it was supposed to do, you could work out how it did it...?

What **do** brains do? What **are** they for?

What **do** brains do? What **are** they for?

Brains are specified by genes that got themselves copied because they tended to produce particular sorts of behaviour – those that led to successful reproduction...

What do brains do? What are they for?

Brains are specified by genes that got themselves copied because they tended to produce particular sorts of behaviour – those that led to successful reproduction...

...and so all brains must have the same single global task (implemented in the choice and execution of domain-specific sub-tasks): to propagate copies of their owners' genes. In other words, they do one thing, and they all do the same thing, and that's what they're for.

What do brains do? What are they for?

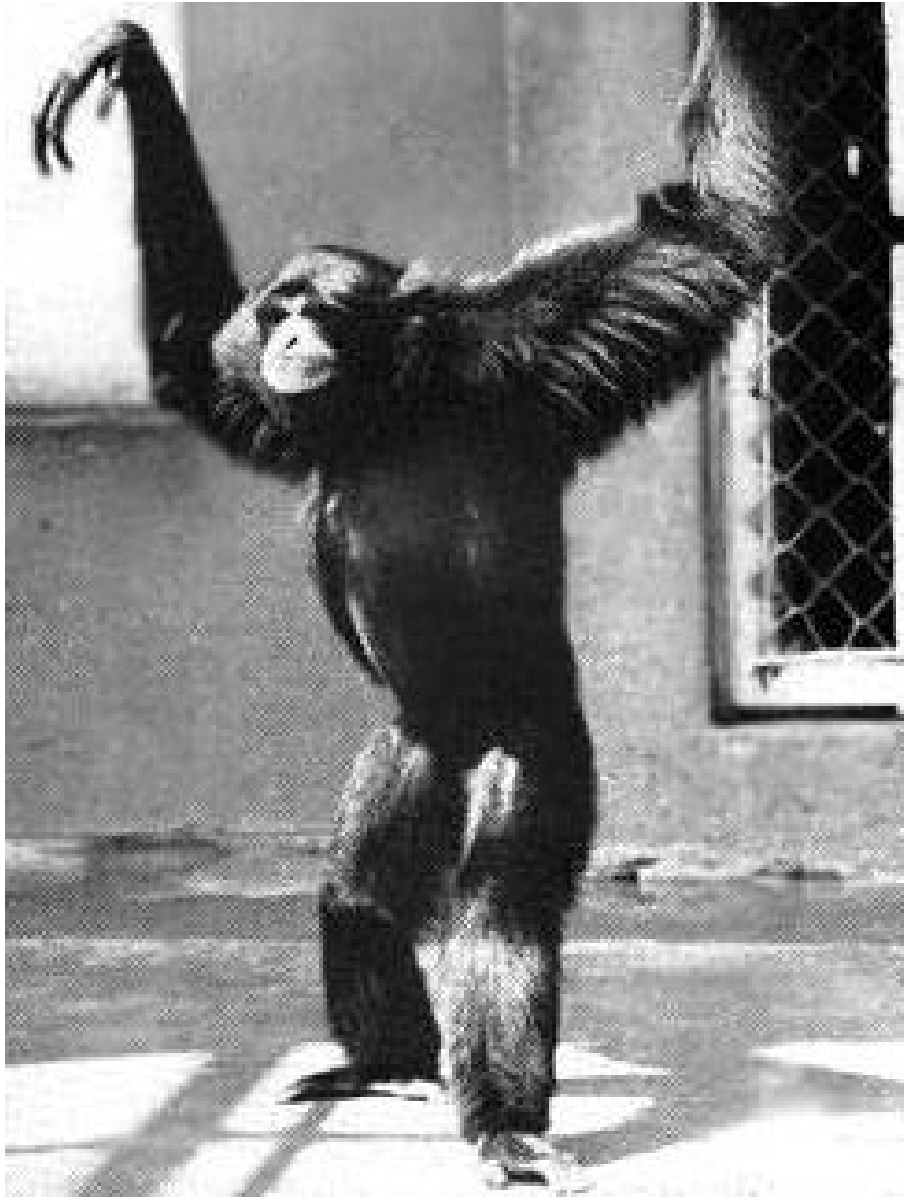
Brains are specified by genes that got themselves copied because they tended to produce particular sorts of behaviour – those that led to successful reproduction...

...and so all brains must have the same single global task (implemented in the choice and execution of domain-specific sub-tasks): to propagate copies of their owners' genes. In other words, they do one thing, and they all do the same thing, and that's what they're for.

Is that it? If evolution always optimises, yes.



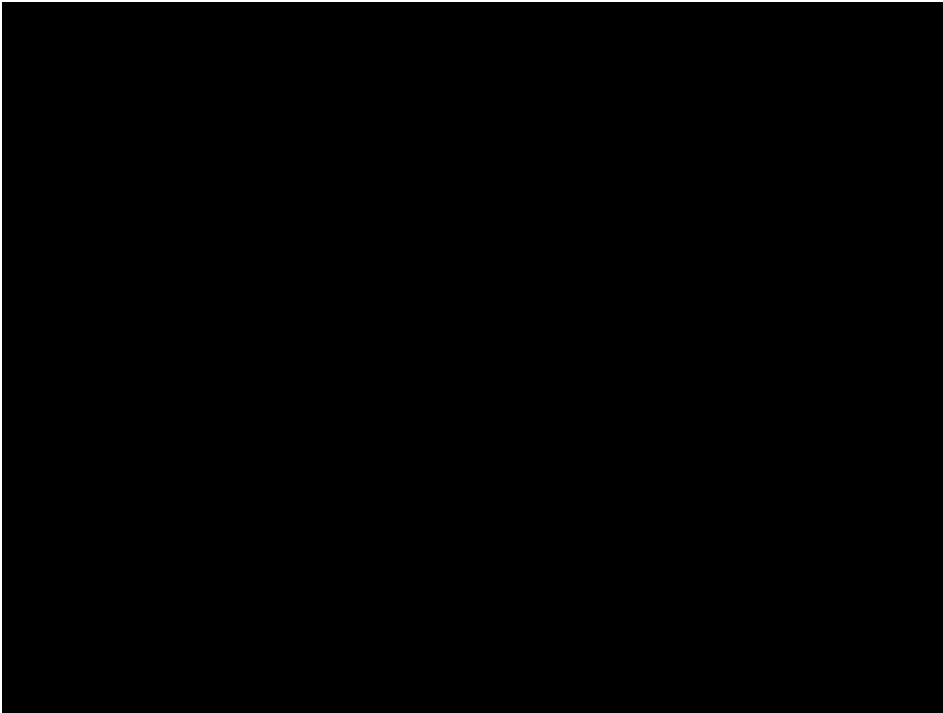
Here's an animal with a design that is probably close to the optimal for living in the topmost branches of trees – the spider monkey.



Here's another tree dweller – the gibbon.

And here's another
– the kangaroo.

Note the tail for
balance, the
relatively powerful
back legs, and the
highly specialised
gait. It is clearly
well adapted to the
vast flat Australian
plains.



And now, meet the tree kangaroo



The Huon tree kangaroo



MDS 2003

So evolution doesn't always optimise...

...but to be fair, the tree kangaroo has evolved claws to grip bark, a more flexible tail, and the ability to move its hind legs separately



Why might evolution fail to optimise?

Not had long enough?

Interference from other processes (e.g. memes)?

Too little variation?

Selection too weak (little competition for tree kangaroos – all are now threatened species)?

Etc. etc.

What are we doing here?

Did we decide that, out of all the possible actions available to us, attending this meeting would maximise the potential for the successful transmission of our genetic material?

What are we doing here?

Did we decide that, out of all the possible actions available to us, attending this meeting would maximise the potential for the successful transmission of our genetic material?

Or was the decision a result of the operation of structures and processes that were evolved in one context, but are now functioning in another?

What is a cognitive system?

View (i): All control systems are cognitive.

“Cognitive processes span the brain, the body, and the environment: to understand cognition is to understand the interplay of all three. Inner reasoning processes are no more essentially cognitive than the skillful execution of coordinated movement or the nature of the environment in which cognition takes place.” (van Gelder and Port, 1995)

What is a cognitive system?

View (ii): Only some control systems – those using internal representations or models in a particular way - are cognitive.

“Cognizers, then, use models (internal and external) in place of directly operating upon the world. Non-cognizers, by contrast, remain trapped in a (potentially very complex and context-variable) web of closed-loop interactions with the very aspects of reality upon which their survival depends.” (Clark and Grush, 1999)

Is this a useful distinction?

Yes.

Higher level control of behaviour (in humans) is mediated by a variety of internal models.

Is this a useful distinction?

Yes.

Higher level control of behaviour (in humans) is mediated by a variety of internal models.

It is not possible for lower levels of control (purely reactive mechanisms) to achieve the same results.

What happens at the borderline?

“For example, if we wish to explore the nature and necessity of the notion of representation in cognitive behavior, then we must examine tasks that are sufficiently “representation-hungry” (Clark & Toribio, 1994). On the other hand, these model agents must be simple enough to be computationally and analytically tractable, so that we have some hope of evolving and analyzing them using techniques that are at most an incremental step beyond what is currently known to be feasible.

The term “minimally cognitive behavior” is meant to connote the simplest behavior that raises cognitively interesting issues.” (Beer 1996)

Examples of ‘minimally cognitive behaviour’

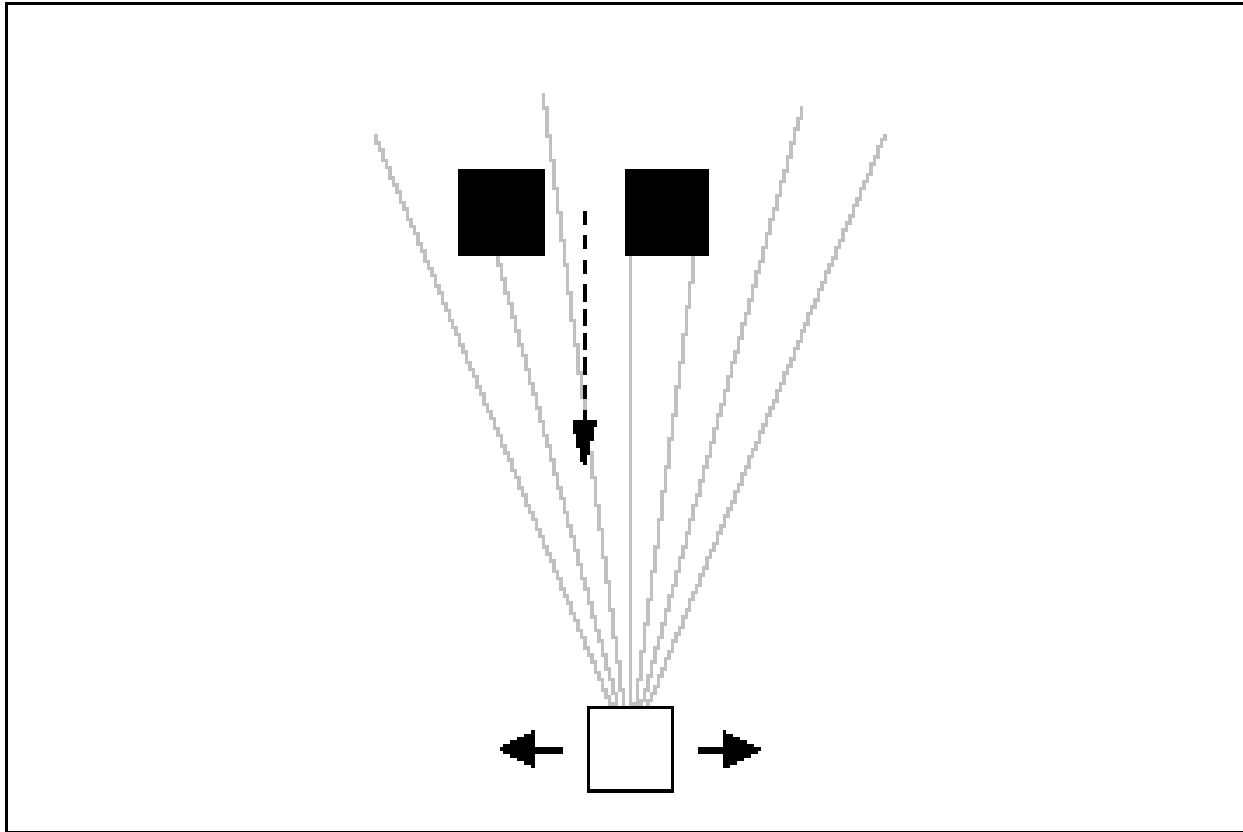
“...agents that can judge the passability of openings relative to their own body size, discriminate between visible parts of themselves and other objects in their environment, predict and remember the future location of objects in order to catch them blind, and switch their attention between multiple distal objects.” (Slocum et al. 2000)

Examples of ‘minimally cognitive behaviour’

“...agents that can judge the passability of openings relative to their own body size, discriminate between visible parts of themselves and other objects in their environment, predict and remember the future location of objects in order to catch them blind, and switch their attention between multiple distal objects.” (Slocum et al. 2000)

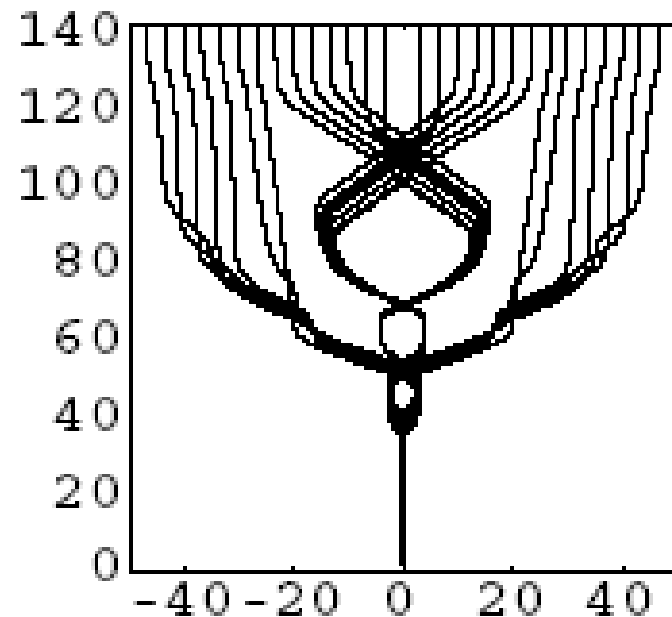
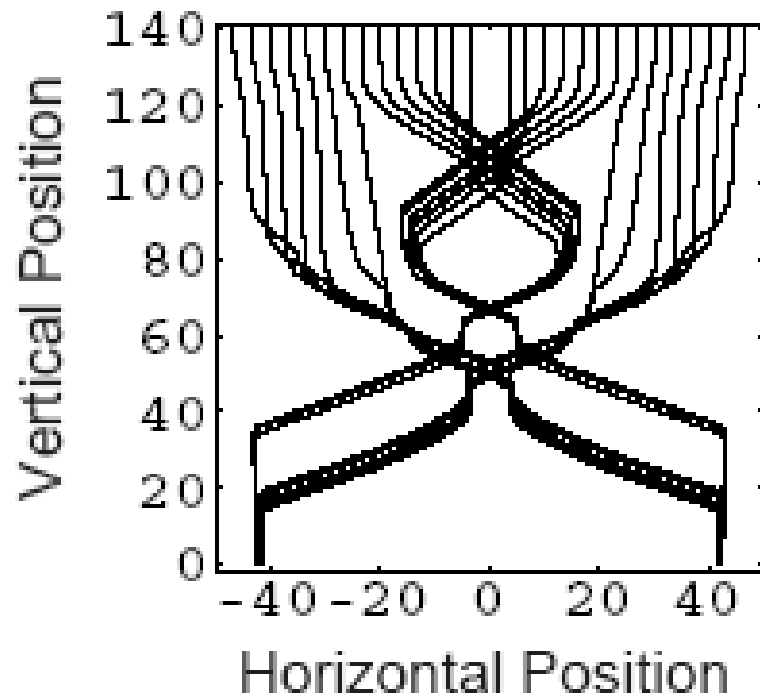
Agents are controlled by simple CTRNNs (continuous time recurrent neural networks) evolved using a simple evolutionary algorithm.

Passability experiments



The grey lines show the agent's sensor array.

Passability experiments



Can a task really be 'representation hungry'?

Suppose you have a walled environment containing lots of similar small objects. You have a number of simple robots with the following three behaviours in decreasing order of priority:

- (1) Turn away from another robot or the boundary
- (2) If pushing more than 3 objects, reverse and turn randomly
- (3) Go straight ahead

- (1) Turn away from another robot or the boundary
- (2) If pushing more than 3 objects, reverse and turn randomly
- (3) Go straight ahead

The robots are to move all the objects until they are in a single pile.

What additional cognitive abilities will they need to be able to do this?

- (1) Turn away from another robot or the boundary
- (2) If pushing more than 3 objects, reverse and turn randomly
- (3) Go straight ahead

The robots are to move all the objects until they are in a single pile.

What additional cognitive abilities will they need to be able to do this?

Let's see what happens if we run the robots as they are...

From local actions to global tasks: Stigmergy and collective robotics

by

Ralph Beekers

ZIF-University of Bielefeld - Free University Brussels

Owen Holland

University of the West of England - Bristol

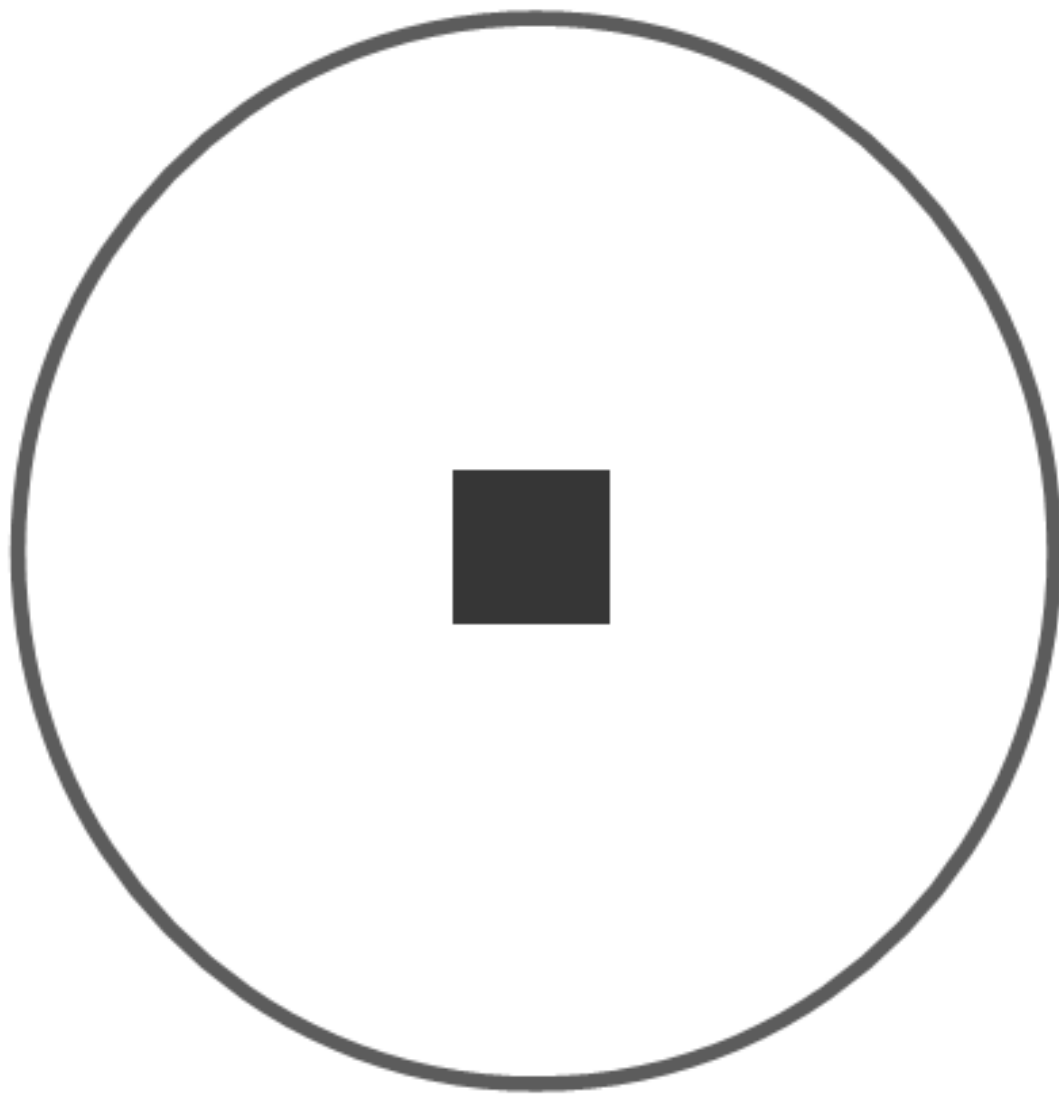
Jean-Louis Deneubourg

Free University Brussels (ULB)

Let's consider the problems of an autonomous embodied agent (an animal or robot)...



Let's consider the problems of an autonomous embodied agent (an animal or robot) in a complex, occasionally novel, dynamic, and hostile world...



Let's consider the problems of an autonomous embodied agent (an animal or robot) in a complex, occasionally novel, dynamic, and hostile world, in which it has to achieve some task (or mission).

How could the agent achieve its task (or mission)?

- by being preprogrammed for every possible contingency?

How could the agent achieve its task (or mission)?

- by being preprogrammed for every possible contingency? No

How could the agent achieve its task (or mission)?

- by being preprogrammed for every possible contingency? No

- by having learned the consequences for the achievement of the mission of every possible action in every contingency?

How could the agent achieve its task (or mission)?

- by being preprogrammed for every possible contingency? No

- by having learned the consequences for the achievement of the mission of every possible action in every contingency? No

How could the agent achieve its task (or mission)?

- by being preprogrammed for every possible contingency? No

- by having learned the consequences for the achievement of the mission of every possible action in every contingency? No

- by having learned enough to be able to predict the consequences of tried and untried actions, by being able to evaluate those consequences for their likely contribution to the mission, and by selecting a relatively good course of action?

How could the agent achieve its task (or mission)?

- by being preprogrammed for every possible contingency? No

- by having learned the consequences for the achievement of the mission of every possible action in every contingency? No

- by having learned enough to be able to predict the consequences of tried and untried actions, by being able to evaluate those consequences for their likely contribution to the mission, and by selecting a relatively good course of action? Maybe

But how could it predict?

But how could it predict?

It might be able to do it directly - somehow

But how could it predict?

It might be able to do it directly - somehow

Or it might be able to do it indirectly, by running some kind of simulation of its actions in the world, enabling it to predict their effects

Here's how Richard Dawkins puts it:

“Survival machines that can simulate the future are one jump ahead of survival machines who can only learn on the basis of overt trial and error.”

Dawkins, 1976

Is it just Dawkins?

No. The idea that some survival machines (animals) runs simulations of actions in the world in order to predict what will happen is quite widespread - e.g. Dennett and Metzinger have written about it.

Some neuroscientists are gathering evidence for it - for example, Germund Hesslow.

Hesslow's 'simulation hypothesis'

"1) *Simulation of actions*. We can activate pre-motor areas in the frontal lobes in a way that resembles activity during a normal action but does not cause any overt movement.

Hesslow's 'simulation hypothesis'

“1) *Simulation of actions*. We can activate pre-motor areas in the frontal lobes in a way that resembles activity during a normal action but does not cause any overt movement.

2) *Simulation of perception*. Imagining that one perceives something is essentially the same as actually perceiving it, but the perceptual activity is generated by the brain itself rather than by external stimuli.

Hesslow's 'simulation hypothesis'

“1) *Simulation of actions*. We can activate pre-motor areas in the frontal lobes in a way that resembles activity during a normal action but does not cause any overt movement.

2) *Simulation of perception*. Imagining that one perceives something is essentially the same as actually perceiving it, but the perceptual activity is generated by the brain itself rather than by external stimuli.

3) *Anticipation*. There are associative mechanisms that enable both behavioural and perceptual activity to elicit other perceptual activity in the sensory areas of the brain. Most importantly, a simulated action can elicit perceptual activity that resembles the activity *that would have occurred if the action had actually been performed.*” (Hesslow 2002)

Embodied cognition

Cognition seems to be very closely linked to embodiment.

This could be because the structures used to represent 'the distal, absent, and non-existent' are the same structures used to deal with real sensing and real motor activity. This is becoming the dominant hypothesis in theories of embodied cognition (e.g. Cruse 2002, Ziemke 2003)

Two questions:

What exactly has to be simulated?

What is needed for simulation?

What exactly has to be simulated?

Whatever affects the mission. In an embodied agent, the agent can only affect the world through the actions of its body in and on the world, and the world can only affect the mission by affecting the agent's body.

What exactly has to be simulated?

Whatever affects the mission. In an embodied agent, the agent can only affect the world through the actions of its body in and on the world, and the world can only affect the mission by affecting the agent's body.

So it needs to simulate those aspects of its body that affect the world in ways that affect the mission, along with those aspects of the world that affect the body in ways that affect the mission.

What exactly has to be simulated?

How does the body affect the world? To some extent through its passive properties, but mainly by being moved through and exerting force on the world, with appropriate speed and accuracy.

What exactly has to be simulated?

How does the body affect the world? To some extent through its passive properties, but mainly by being moved through and exerting force on the world, with appropriate speed and accuracy.

How does the world affect the body? Through the spatially distributed environment (through which the body must move) and through the properties of the objects in it (cf. food, predators, poisons, prey, competitors, falling coconuts, etc. for animals)

What is needed for simulation?

Some structure or process corresponding to a state of the world that, when operated on by some process or structure corresponding to an action, yields an outcome corresponding to and interpretable as the consequences of that action.

What is needed for simulation?

I like to call these structures or processes 'internal models', because they are like working models rather than static representations, and because the term was used in this sense by Craik, and later by Johnson-Laird and others.

What is needed for simulation?

I like to call these structures or processes ‘internal models’, because they are like working models rather than static representations, and because the term was used in this sense by Craik, and later by Johnson-Laird and others.

So we require a model (or linked set of models) that includes the body, and how it is controlled, and the spatial aspects of the world, and the (kinds of) objects in the world, and their spatial arrangement. But consider...

What is needed for simulation?

The body is always present and available, and changes slowly, if at all. When it moves, it is usually because it has been commanded to move.

What is needed for simulation?

The body is always present and available, and changes slowly, if at all. When it moves, it is usually because it has been commanded to move.

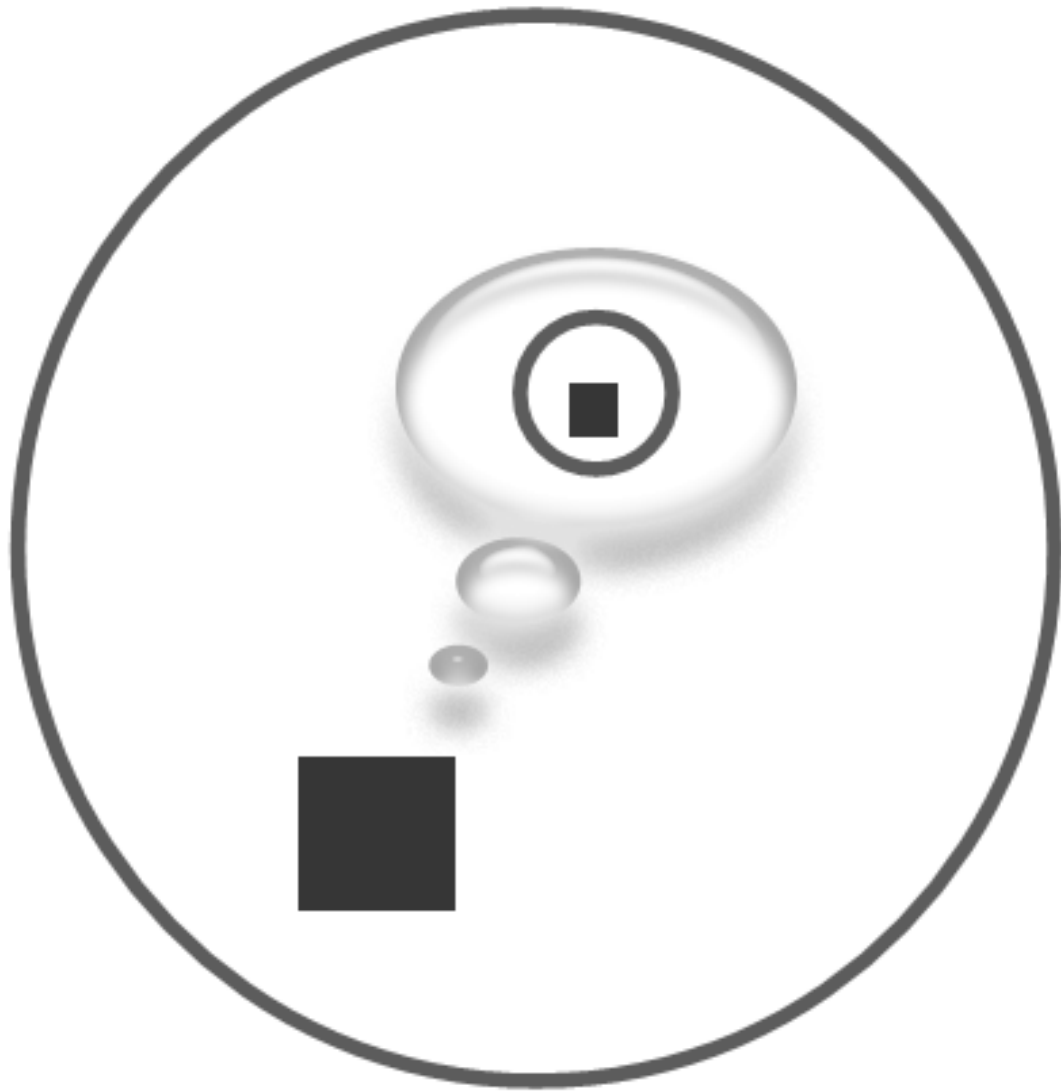
The world is different. It is 'complex, occasionally novel, dynamic, and hostile'. It's only locally available, and may contain objects of known and unknown kinds in known and unknown places.

What is needed for simulation?

The body is always present and available, and changes slowly, if at all. When it moves, it is usually because it has been commanded to move.

The world is different. It is 'complex, occasionally novel, dynamic, and hostile'. It's only locally available, and may contain objects of known and unknown kinds in known and unknown places.

How should all this be modelled? As a single model containing body, environment, and objects?

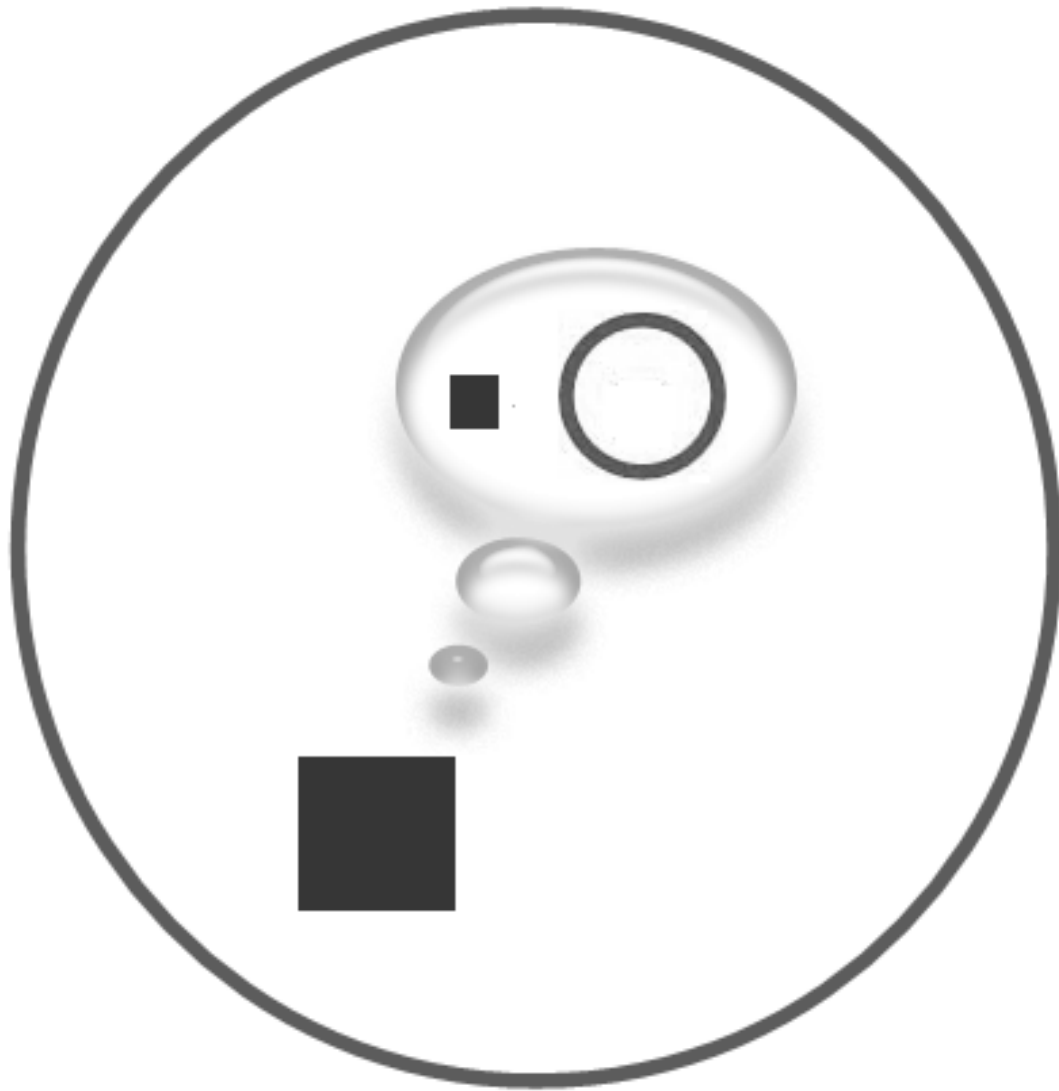


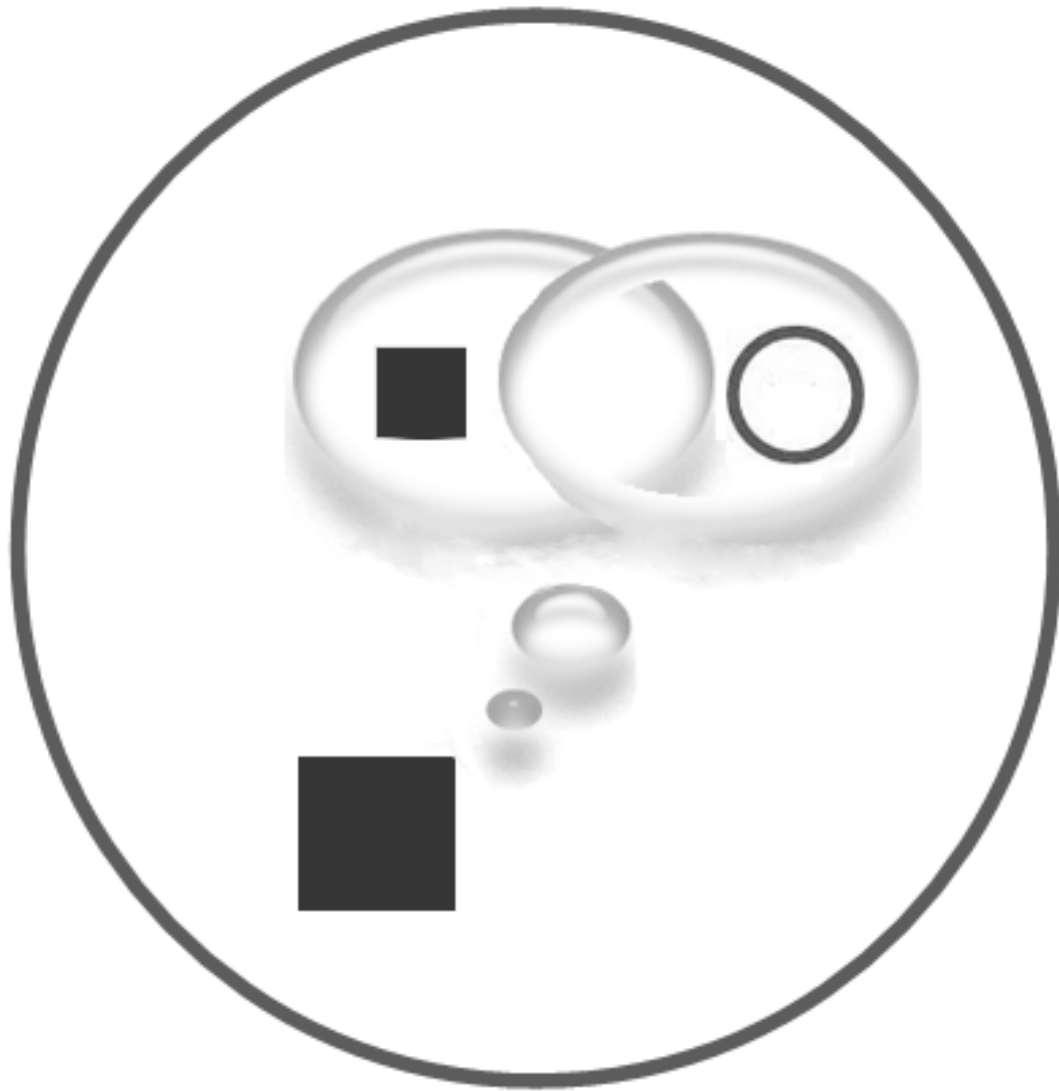
What is needed for simulation?

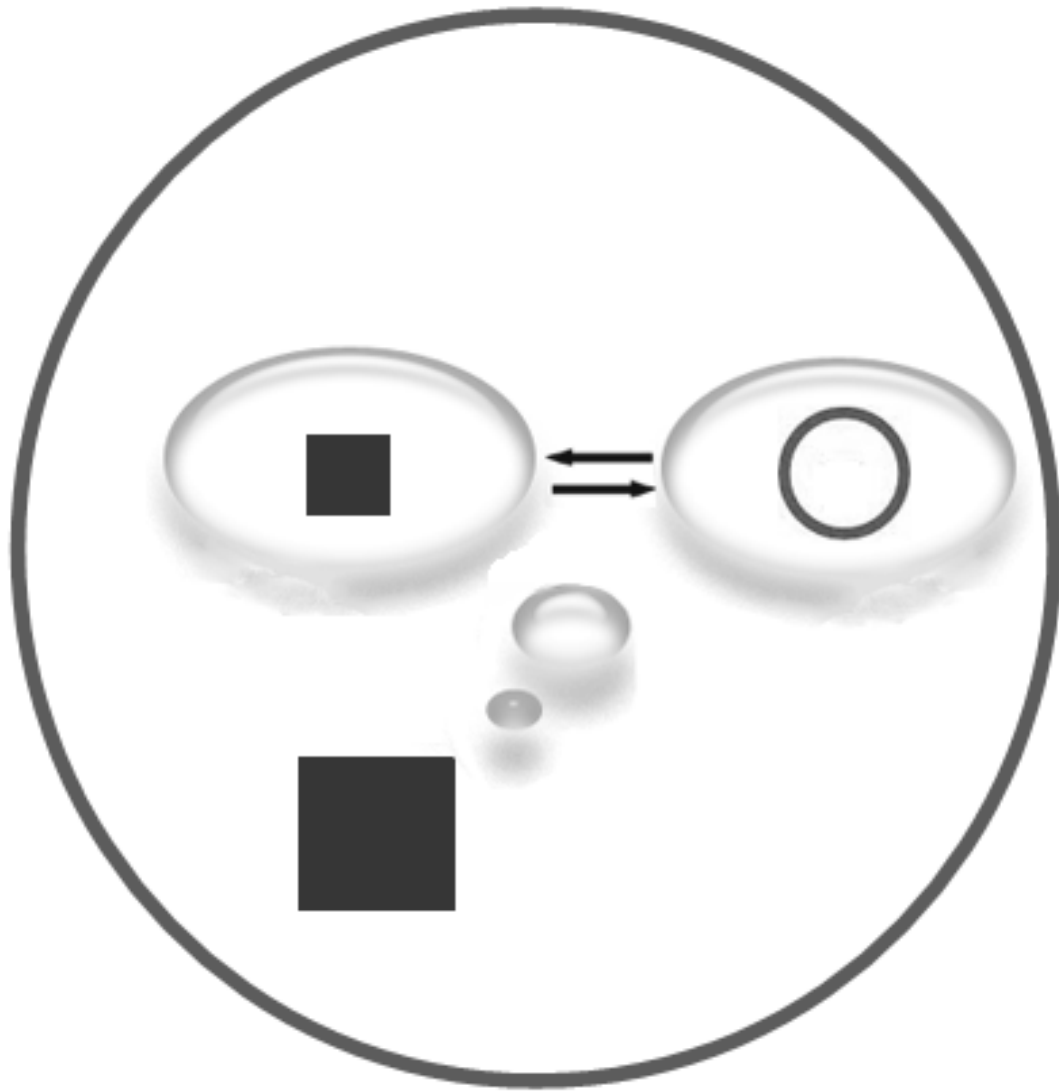
The body is always present and available, and changes slowly, if at all. When it moves, it is usually because it has been commanded to move.

The world is different. It is 'complex, occasionally novel, dynamic, and hostile'. It's only locally available, and may contain objects of known and unknown kinds in known and unknown places.

How should all this be modelled? As a single model containing body, environment, and objects? Or as a separate model of the body coupled to and interacting with the other modelled components?







What happens in the human animal?

What happens in the human animal?

“...(I)t is always obvious to you that there are some things you can do and others you cannot given the constraints of your body and of the external world. (You know you can't lift a truck...) Somewhere in your brain there are representations of all these possibilities, and the systems that plan commands...need to be aware of this distinction between things they can and cannot command you to do....To achieve all this, I need to have in my brain not only a representation of the world and various objects in it but also a representation of myself, including my own body within that representation....In addition, the representation of the external object has to interact with my self-representation....” (Ramachandran and Blakeslee 1998).

Does the brain model the body?

Yes, in many ways. It models the muscular control of movement, using forward models and inverse models (Ito, Kawato, Wolpert etc.)

It also predicts the nature and timing of the internal and external sensory inputs that will be produced if the movement is executed correctly (Frith, Blakemore). This is useful because feedback is too slow to guide rapid movements, and such prediction allows early correction.

Does the brain model the body?

Ramachandran and Blakeslee describe a host of body image phenomena involving phantom limbs. In one case, a patient with congenital absence of both arms had apparently 'normal' phantom limbs from an early age. Some components of the internal model of the body may be innate.

Does the brain model the world?

Yes, in many ways. It models space, and it models the nature and behaviour of objects, and much of this modelling is innate.

Useful reading (for me anyway): *Wild Minds*, by Marc Hauser.

Does the brain model the world?

Yes, in many ways. It models space, and it models the nature and behaviour of objects, and much of this modelling is innate.

Useful reading (for me anyway): *Wild Minds*, by Marc Hauser.

The use of forward and inverse models to predict external sensory inputs from movements doesn't just allow improved motor control - it gives rise to a compelling theory of internal representation of the external world - Rick Grush's emulation theory, solidly rooted in control engineering.

Exactly how can simulation help the agent?

All simulation can tell you is what will probably happen if you do the action Z , or the action sequence XYZ .

Exactly how can simulation help the agent?

All simulation can tell you is what will probably happen if you do the action Z , or the action sequence XYZ .

(a) If the outcome is a desired state Z^* (the 'goal') then the action (sequence) can easily be triggered. Search and simulation alone are enough.

Exactly how can simulation help the agent?

All simulation can tell you is what will probably happen if you do the action Z , or the action sequence XYZ .

(a) If the outcome is a desired state Z^* (the 'goal') then the action (sequence) can easily be triggered. Search and simulation alone are enough.

(b) If the outcome is a state Z^* which is then evaluated for its likely contribution to the mission, the action (sequence) may be selected, or preferred over others once they have been evaluated. Search and simulation alone are **not** enough - you also need evaluation, storage, retrieval etc.

Exactly how can simulation help the agent?

All simulation can tell you is what will probably happen if you do the action Z , or the action sequence XYZ .

(a) If the outcome is a desired state Z^* (the 'goal') then the action (sequence) can easily be triggered. Search and simulation alone are enough.

(b) If the outcome is a state Z^* which is then evaluated for its likely contribution to the mission, the action (sequence) may be selected, or preferred over others once they have been evaluated. Search and simulation alone are not enough - you also need evaluation, storage, retrieval etc.

MUCH MORE COMPLEX

Evaluation

“Most recordings from neurons in the monkey cortex have focused on the abstract sensory information that they encode, or the actions that they elicit. Only recently has it been discovered that many cortical neurons also respond to properties of the reward....These results suggest that **a major function of the cerebral cortex has been ignored**, and in particular that the cortex may be as concerned with representing reward contingencies as it is to representing properties of the physical world.”

Brains, Rewards, and Game Theory Workshop (2003)

Swartz Center for Computational Neuroscience, UCSD

Is all this worth it?

You only have to do better than you would do if you didn't simulate and evaluate (and didn't pay the time, energy, capital, development, and running costs of simulation and evaluation). You don't have to calculate utility perfectly. You don't have to search all possibilities, or search with maximum efficiency. And it has to be quick.

**BUT IF YOU CAN DO THIS YOU WILL BEAT A
PURELY REACTIVE NON-COGNITIVE SYSTEM**

Conclusions

Perhaps...

We should pay more attention to what the whole brain is for, and less to what the parts (can) do

Conclusions

Perhaps...

We should pay more attention to what the whole brain is for, and less to what the parts (can) do

We should distinguish between cognitive and non-cognitive systems, and pay more attention to the cognitive ones

Conclusions

Perhaps...

We should pay more attention to what the whole brain is for, and less to what the parts (can) do

We should distinguish between cognitive and non-cognitive systems, and pay more attention to the cognitive ones

We should recognise that we ourselves are truly cognitive systems

Consciousness

‘...I must present a theory (of sentience) that addresses questions like these: If we could ever duplicate the information processing in the human mind as an enormous computer program, would a computer running the program be conscious?...etc...etc...’

Steven Pinker 1997

Consciousness

‘...I must present a theory (of sentience) that addresses questions like these: If we could ever duplicate the information processing in the human mind as an enormous computer program, would a computer running the program be conscious?...etc...etc...’

Beats the heck out of me! I have some prejudices, but no idea of how to look for a defensible answer. And neither does anyone else.’

Steven Pinker 1997

What is consciousness?

‘Consciousness is a puzzler.’

Charles Darwin

But all this has been merely cognitive. What has it to do with consciousness?

What Dawkins (1976) said next:

“Survival machines that can simulate the future are one jump ahead of survival machines who can only learn on the basis of overt trial and error. ..The evolution of the capacity to simulate seems to have culminated in subjective consciousness...Perhaps consciousness arises when the brain’s simulation of the world becomes so complete that it must include a model of itself.”

But all this has been merely cognitive. What has it to do with consciousness?

What Dawkins (1976) said next:

“Survival machines that can simulate the future are one jump ahead of survival machines who can only learn on the basis of overt trial and error. ..The evolution of the capacity to simulate seems to have culminated in subjective consciousness...Perhaps consciousness arises when the brain’s simulation of the world becomes so complete that it must include a model of itself.”

How about ‘...a model of the whole machine’?

"...consciousness requires that the brain must represent not just the object, not just a basic self structure, ***but the interaction of the two***....This is still an atypical foundation for a theory of consciousness, given that until recently, it was implicitly assumed that the self could be left out of the equation. There has been a recent sea change on this crucial point..."

Douglas Watt 2000, review of Damasio's "The Feeling of What Happens" (Damasio 1999).

In other words...

Intelligent cognitive behaviour in an embodied agent may depend on the possession and manipulation of an internal model of the agent (the IAM) interacting with an internal model of the world

AND

the presence and interaction of these models may also underlie the production of consciousness.

And there's more...

"The phenomenal self is a virtual agent perceiving virtual objects in a virtual world...I think that 'virtual reality' is the best technological metaphor which is currently available as a source for generating new theoretical intuitions ...heuristically the most interesting concept may be that of 'full immersion'."
(Metzinger 2000)

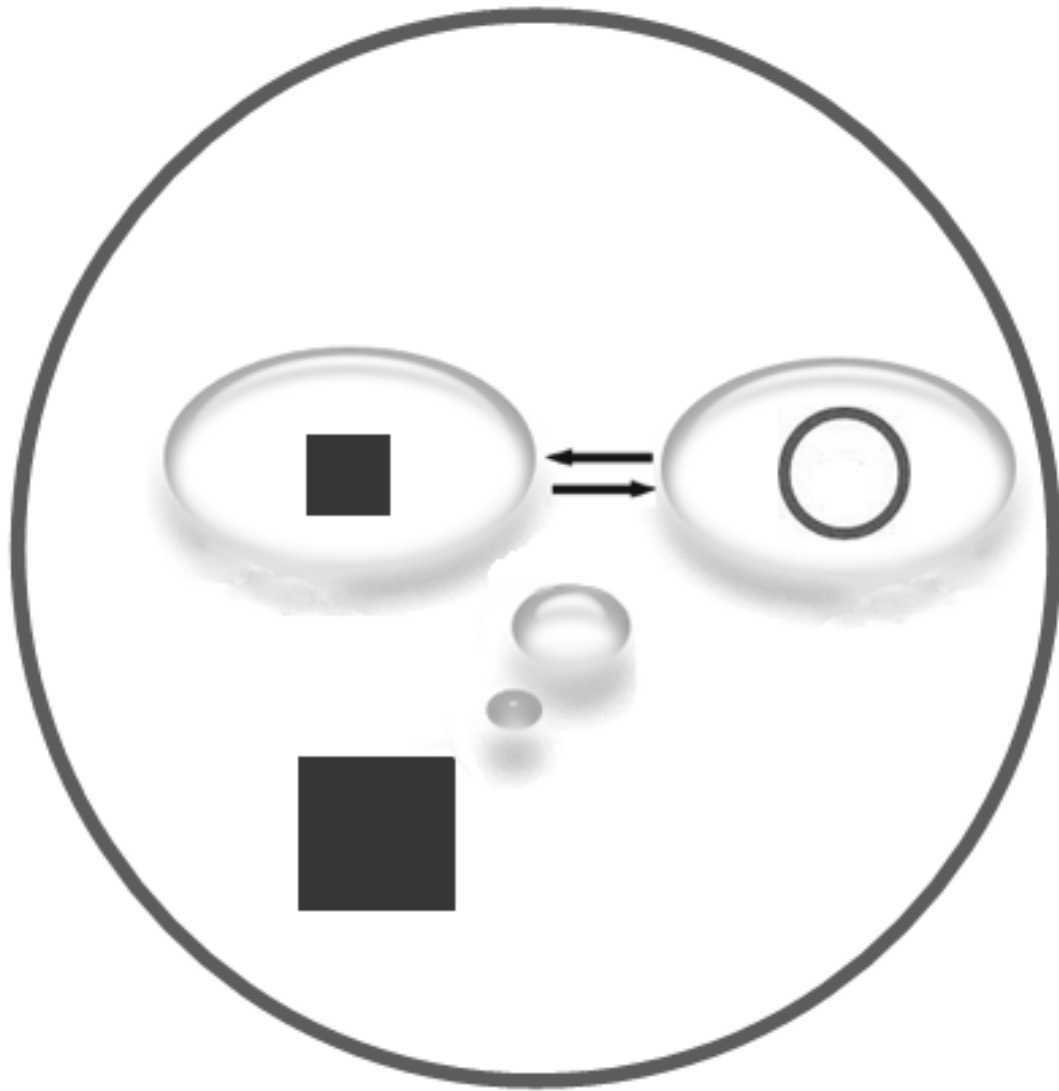
And there's more...

The phenomenal self-model "...is a plastic multimodal structure that is plausibly based on an innate and 'hardwired' model of the spatial properties of the system (e.g. a 'long-term body image'...) while being functionally rooted in elementary bioregulatory processes...."

(Metzinger 2000)

A hypothesis

In humans (and some animals?) it is the internal agent model that is conscious, not the agent itself.



And...

In order to produce accurate predictions, the agent model and the world model must be constantly updated with changes in the agent and the world that affect the mission, whether planning is currently taking place or not.

And so...

In order to produce accurate predictions, the agent model and the world model must be constantly updated with changes in the agent and the world that affect the mission, whether planning is currently taking place or not. The 'contents' of consciousness are the effects on the internal agent model of its own dynamics, of direct updates, and also of updates to the world model to which the IAM is coupled.

And so...

In order to produce accurate predictions, the agent model and the world model must be constantly updated with changes in the agent and the world that affect the mission, whether planning is currently taking place or not. The 'contents' of consciousness are the effects on the internal agent model of its own dynamics, of direct updates, and also of updates to the world model to which the IAM is coupled. The IAM does not control the body, but attributes the updates of bodily movements to its own agency.

And so...

In order to produce accurate predictions, the agent model and the world model must be constantly updated with changes in the agent and the world that affect the mission, whether planning is currently taking place or not. The 'contents' of consciousness are the effects on the internal agent model of its own dynamics, of direct updates, and also of updates to the world model to which the IAM is coupled. The IAM does not control the body, but attributes the updates of bodily movements to its own agency. The peculiarities of consciousness are simply the natural characteristics of such a system.

A proposal

The way to study these phenomena is to build a suitably complex robot, to embed it in a suitably complex environment and to examine the robot's behaviour and internal processes as it learns to cope with its mission.

A funded proposal!

The way to study these phenomena is to build a suitably complex robot, to embed it in a suitably complex environment and to examine the robot's behaviour and internal processes as it learns to cope with its mission.

£493,000 (\$907,000) from the Engineering and Physical Science Research Council

Start date April 1st 2004, duration 3 years.

Team: Owen Holland, David Gamez, Rob Knight (Computer Science, Essex); Tom Troscianko, Iain Gilchrist, Ben Vincent (Psychology, Bristol)

How will we know if it's conscious?

I don't know. But people are beginning to devise some useful frameworks for answering the question.

How will we know if it's conscious?

I don't know. But people are beginning to devise some useful frameworks for answering the question.

Igor Aleksander has proposed 5 axioms:

AXIOM 1: A SENSE OF PLACE We feel that we are at the centre of an "out there" world, and we have the ability to place ourselves in the world around us.

AXIOM 2: IMAGINATION We can 'see' things that we have experienced in the past, and we can also conjure up things we have never seen. Reading a novel can conjure up mental images of different worlds, for example.

AXIOM 3: DIRECTED ATTENTION Our thoughts are not just passive reflections of what is happening in the world - we are able to focus our attention, and we are conscious only of that to which we attend.

AXIOM 4: PLANNING We have the ability to carry out "what if" exercises. Scenarios of future events and actions can be mapped out in our minds even if we are just sitting still.

AXIOM 5: DECISION/EMOTION Emotions guide us into recognising what is good for us and what is bad for us, and in acting accordingly.

How will we know if it's conscious?

I don't know. But people are beginning to devise some useful frameworks for answering the question.

Thomas Metzinger has identified 11 constraints on “...what makes a neural representation a phenomenal representation”

(T Metzinger, 2003: Being No-one: the self-model theory of subjectivity. 699 pages!)

Metzinger's 11 constraints

- (1) Global availability
- (2) Activation within a window of presence
- (3) Integration into a coherent global state
- (4) Convolved holism
- (5) Dynamicity
- (6) Perspectivalness
- (7) Transparency
- (8) Offline activation
- (9) Representation of intensities
- (10) "Ultrasmoothness": Homogeneity of simple content
- (11) Adaptivity

Metzinger's final constraint

"Suffering starts on the level of Phenomenal Self Models. You cannot consciously suffer without having a globally available self-model. The PSM is the decisive neurocomputational instrument not only in developing a host of new cognitive and social skills but also in forcing any strongly conscious system to functionally and representationally appropriate its own disintegration, its own failures and internal conflicts. Phenomenal appropriation goes along with functional appropriation."

Metzinger's final constraint

“Evolution is not only marvellously efficient but also ruthless and cruel to the individual organism. Pain and any other nonphysical kind of suffering, generally any representational state characterized by a "negative valence" and integrated into the PSM are now phenomenally owned. Now it inevitably, and transparently, is my own suffering. The melodrama, but also the potential tragedy of the ego both start on the level of transparent self-modeling. **Therefore, we should ban all attempts to create (or even risk the creation of) artificial and postbiotic PSMs from serious academic research.**”

T. Metzinger, Being No-One (p 622).