Intelligent Control
and Cognitive Systems

brings you...

Consciousness and Cognitive Systems

Joanna J. Bryson
University of Bath, United Kingdom

Consciousness & Cognitive Systems

- Can an artificial cognitive system be conscious?
- Who cares?
- Why care?
- What is consciousness in the first place?

KC Green, *Gun Show*, #513

Cognitive Systems & Philosophy

- Science fiction uses robots and aliens to examine the human condition; the future to examine the present.
- AI does the same thing.
  - ... but, AI is also real.
  - Well, some of it is real.
    - Some of it is tangled with Sci Fi.

Roadmap for Conscious Machines

1. (-1) Disembodied
1. (0) Isolated
1. Decontrolled
2. Reactive
3. Adaptive
4. Attentional
5. Executive
6. Emotional
7. Self-conscious
8. Empathic
9. Social
10. Human-like
11. Super Conscious

Arrabales et al 2009

Roadmap for Conscious Machines

1. (-1) Disembodied
1. (0) Isolated
1. Decontrolled
2. Reactive — Sensing to action: *intelligence*
3. Adaptive
4. Attentional — Unconsciousness is more conscious!

Arrabales et al 2009

Roadmap for Conscious Machines

5. Executive — multiple goals (unconscious 2)
6. Emotional — "human like"???
7. Self-conscious — knows about self
8. Empathic — knowledge (k) of others
9. Social — k of other's k of self
10. Human-like — use Interweb to extend mind
11. Super Conscious — multiple streams!

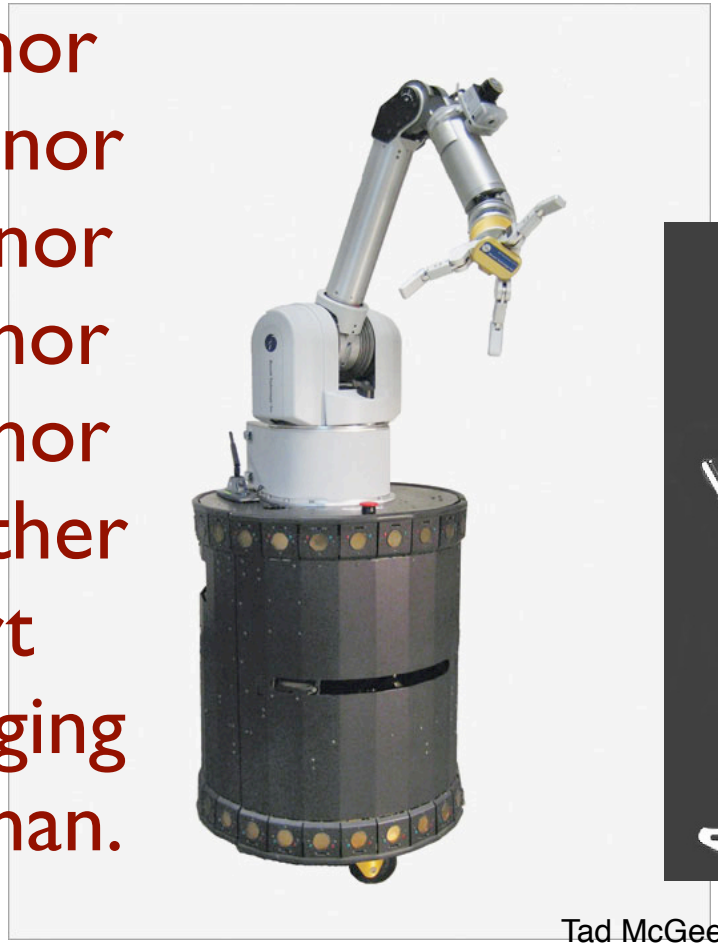
Consciousness ?= Like ME!!!

- From an AI & even Computer Science perspective, many of these criteria are easy to achieve.
  - E.g. perfect self knowledge.
- Consciousness is easy but combinatorics is hard – computational explanation for biological phenomenon of unconsciousness?

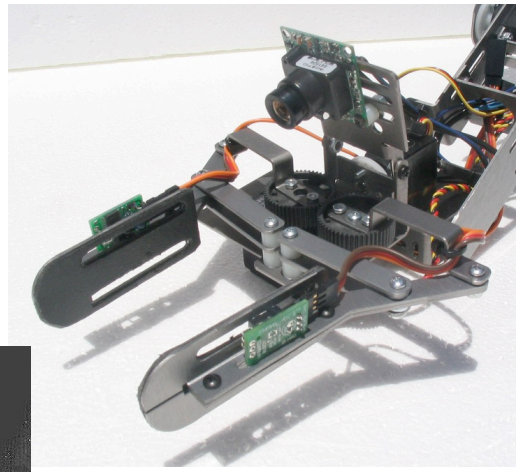
Bryson, Philosophy Magazine, 2007

What's Consciousness?

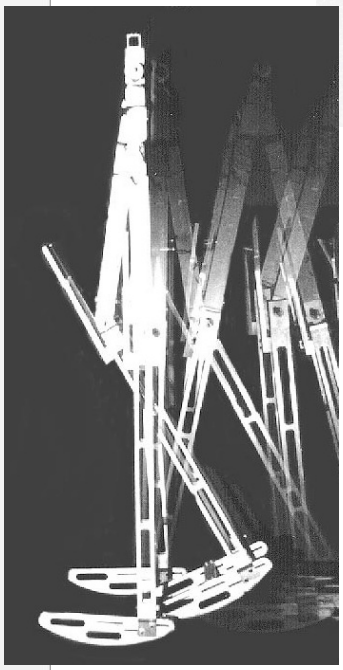
it is nor hand, nor foot, nor arm, nor face, nor any other part belonging to a man.



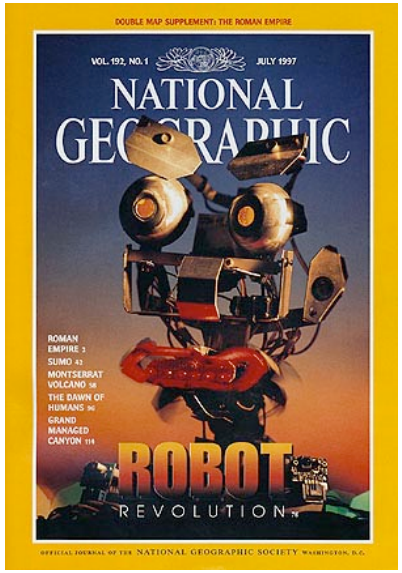
Glenn Matsumura, Wired 2007



SG5-UT Robotic Arm

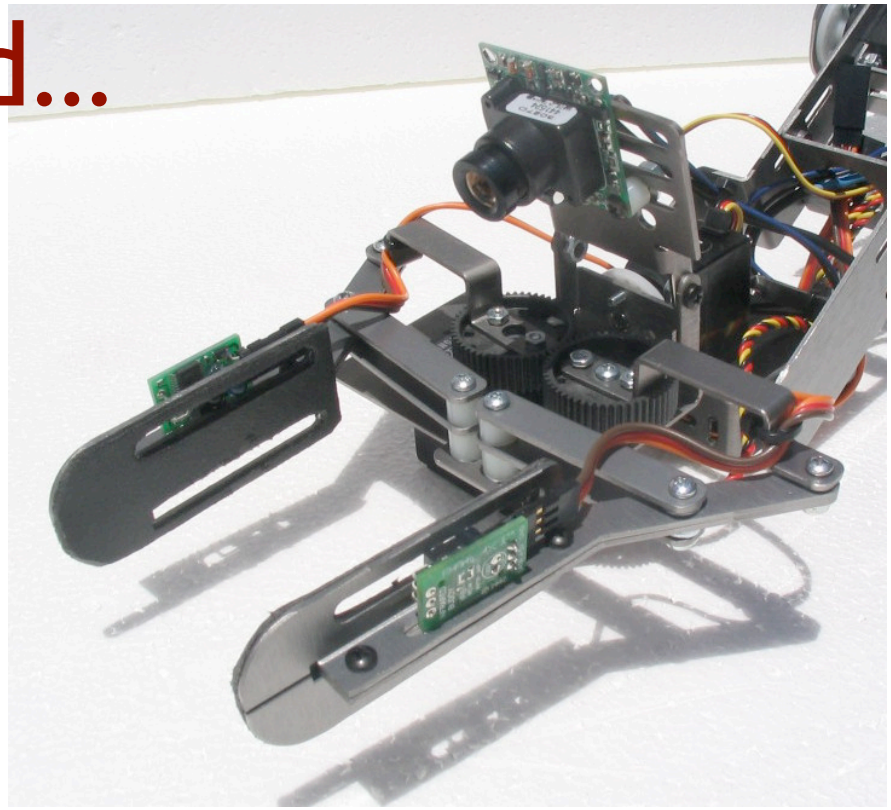


Tad McGeer's passive dynamic walker



Chuck Rosenberg's IT, 1997

If this can be a hand...



...what could a mind be like?

Modelling Natural Intelligence

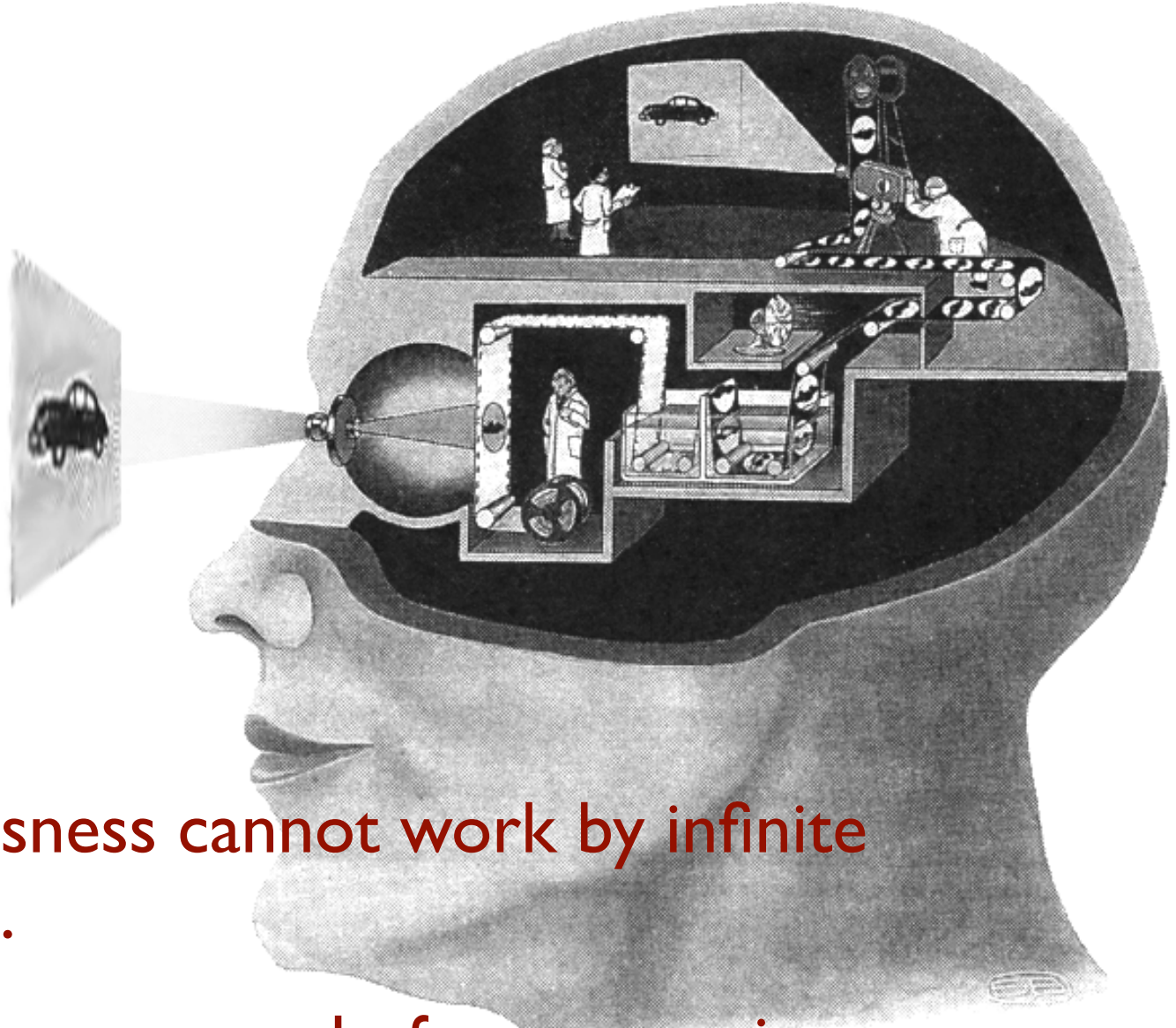
- One of the best ways to understand how something works is to build it yourself.
- AI is used in scientific modelling, but also in Philosophy.

Dennett: "Intuition Pumps"

Consciousness as per Dennett

- The term conscious is itself culturally evolved.
  - May not refer to any one psychological phenomenon.
  - Like light before modern physics.

Dennett vs. The Cartesian Theatre

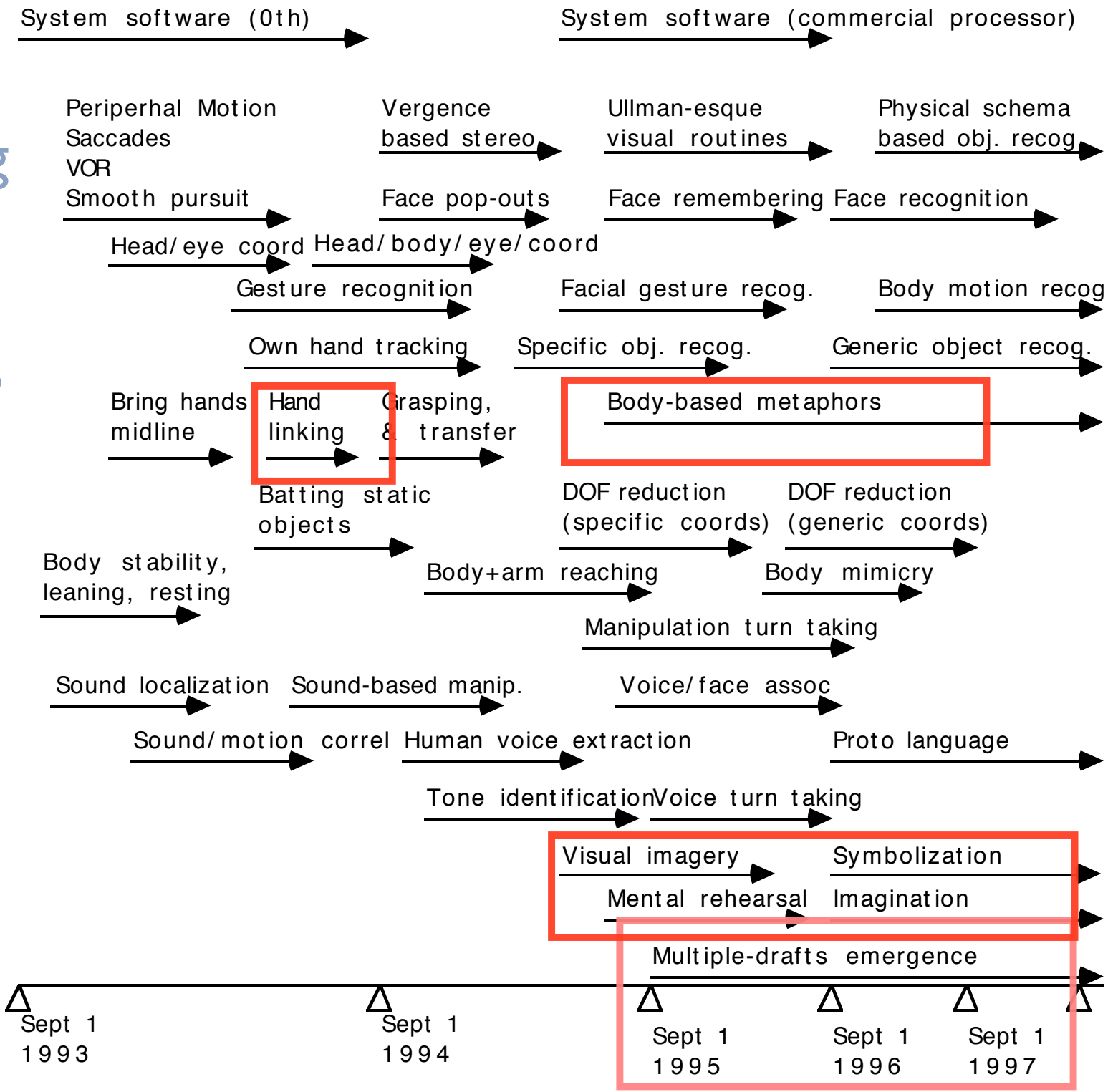


- Consciousness cannot work by infinite recursion.
  - Must be composed of non-conscious elements.
  - Nothing inside you is conscious; you are.

Multiple Drafts / The Attentional Spotlight

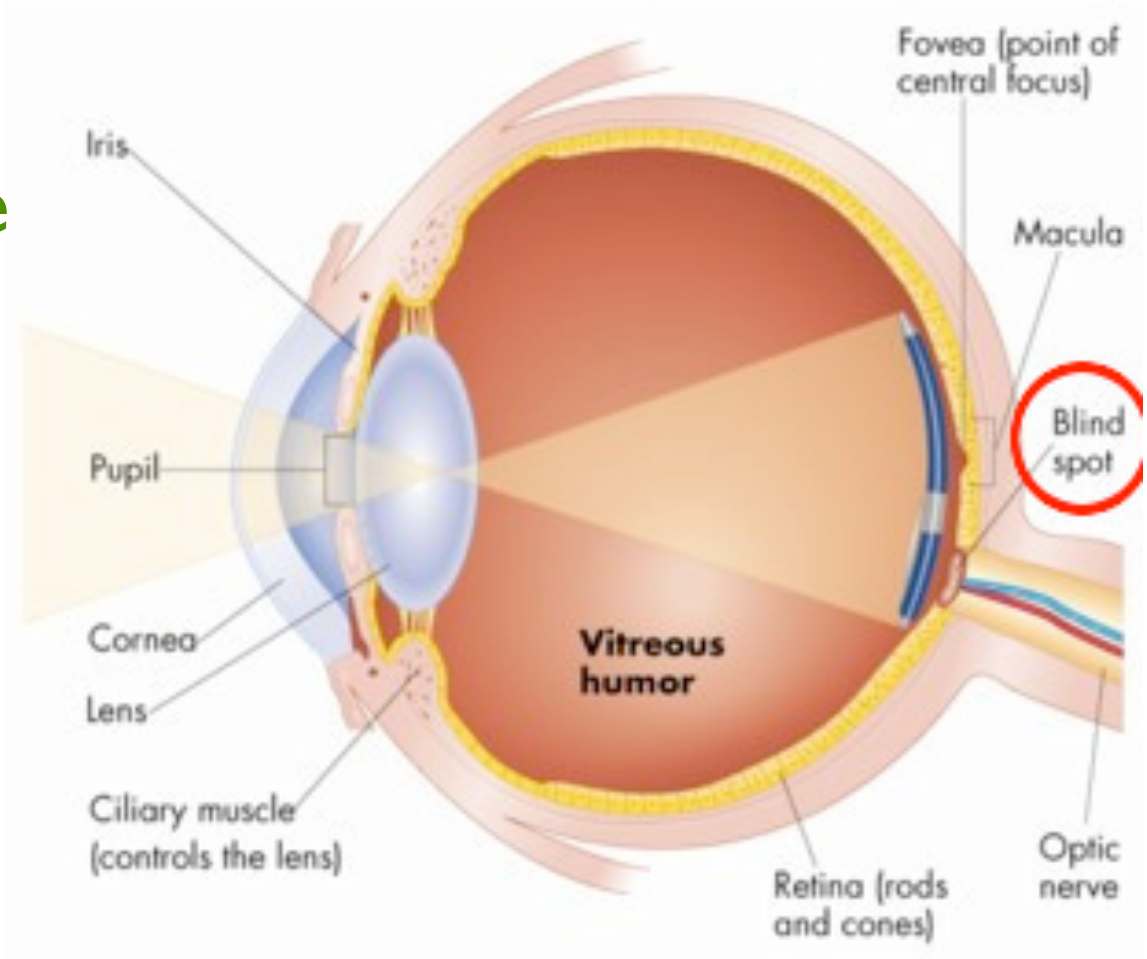
- There are many candidate parallel processes which could be conscious.
  - Only one is – leaves trace in episodic memory.
- Not necessarily determined in order, e.g. if driving may 'see' something you hit only after you hear the bump.
  - ∴ Not really conscious all the time?

"Building Brains for Bodies", Brooks & Stein (1993), MIT AI lab tech report 1439.



Fill In and Confabulation

Things like the driving story & the fact we are never aware of our blind spot unless we really go out of the way to test for it indicate we cannot trust our intuitions about consciousness.



Dennett Critics

- Some people really hate these ideas.
- Chalmers is the main anti-Dennett champion.
  - Chalmers' Hard Problem: Explaining qualia.
    - How do you know someone else sees red the same way you do?

The Zombie Problem (seriously)

- A standard problem in philosophy: how would you tell if someone wasn't conscious?
- Dennett: the zombie idea is incoherent.
  - (likes Brooks, embodiment theory.)
  - Consciousness is what it's like to act human.
  - There's nothing else.
- Critics: Dennett thinks we're all zombies!

Popular Theories of Consciousness

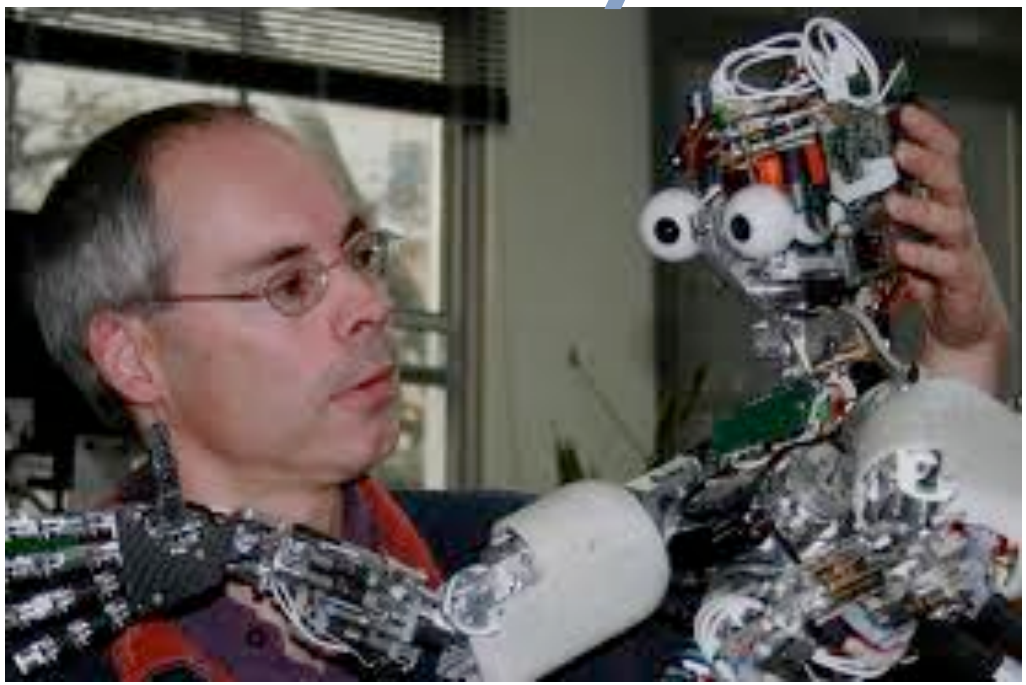
- Consciousness is self-awareness.
- Consciousness requires language.
- Consciousness is the root of ethical obligation / a soul.
- Consciousness is a special pattern of energy (Dahaene)
- Consciousness is a special level of information integration (Tononi)

What People Like in Consciousness Theories

- We'll never understand consciousness.
- We will understand it, but not in 100 years.
- I have a quantitative, scientific measure of consciousness, but it will take 60 years until we can check if I'm right (Tononi).
- Only humans are conscious.

Currently the most popular theory in Cognitive Systems Research is Barr's Global Workspace Theory

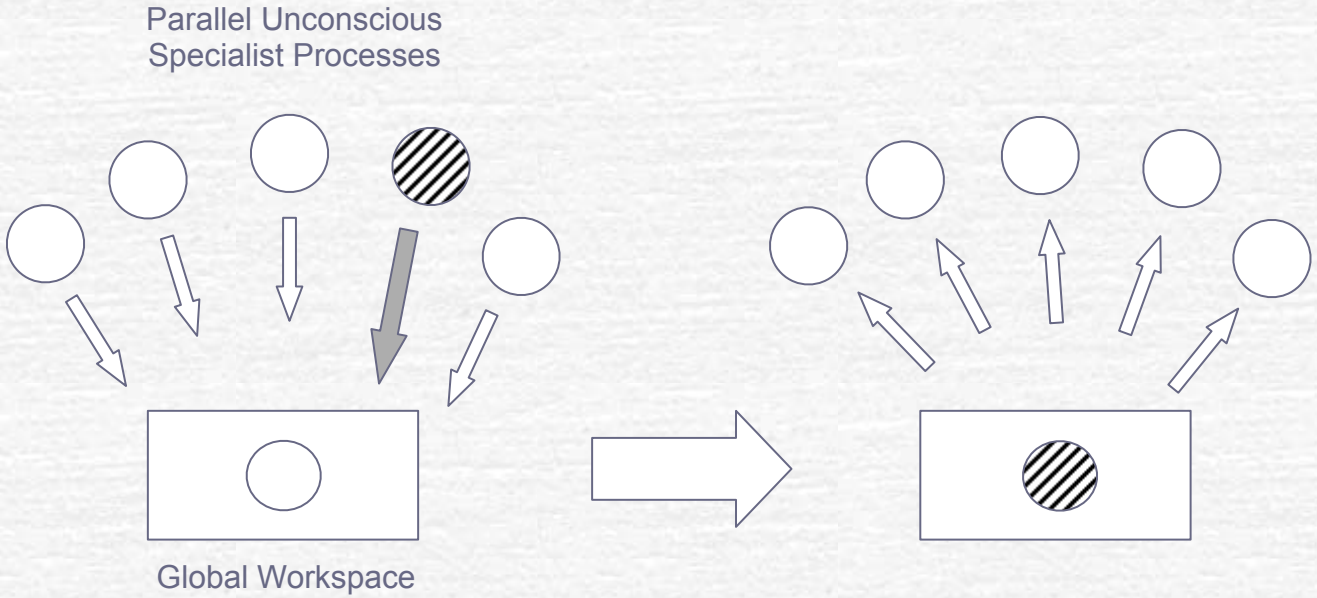
Upcoming slides by
Murray Shanahan



Neural Parallelism

- An animal's nervous system is massively parallel
- Massive parallelism surely underpins human cognitive prowess
- So how are the massively parallel computational resources of an animal's central nervous system harnessed for the benefit of that animal?
- How can they orchestrate a coherent and flexible response to each novel situation?
- Nature has solved this problem. How?

Global Workspace Architecture

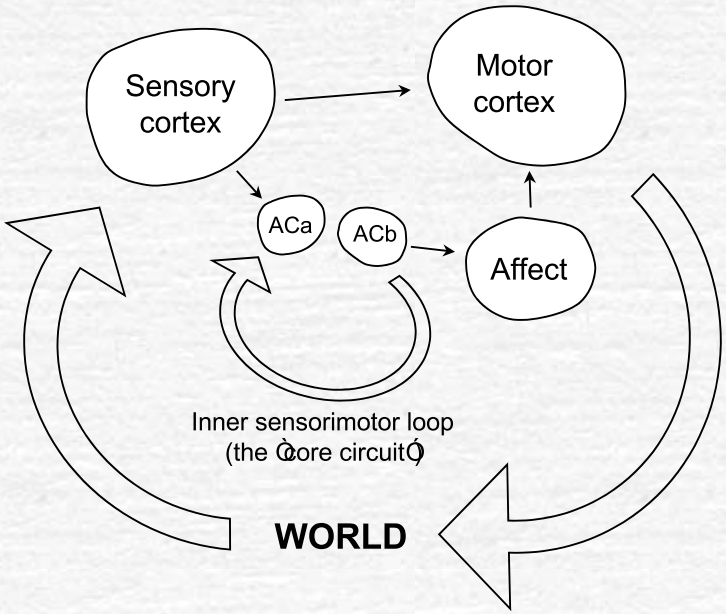


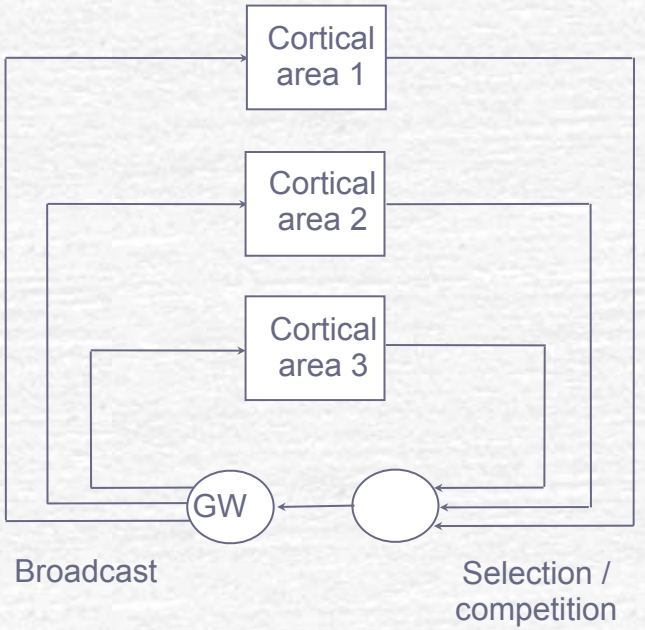
- Multiple parallel *specialist* processes compete and co-operate for access to a *global workspace*
- If granted access to the global workspace, the information a process has to offer is *broadcast* back to the entire set of specialists

Conscious vs Non-Conscious

- Global workspace theory (Baars) hypothesises that the mammalian brain instantiates such an architecture
- It also posits an empirical distinction between conscious and non-conscious information processing
- Information processing in the parallel specialists is non-conscious
- Only information that is broadcast is consciously processed

Combining a GW with Internal Simulation

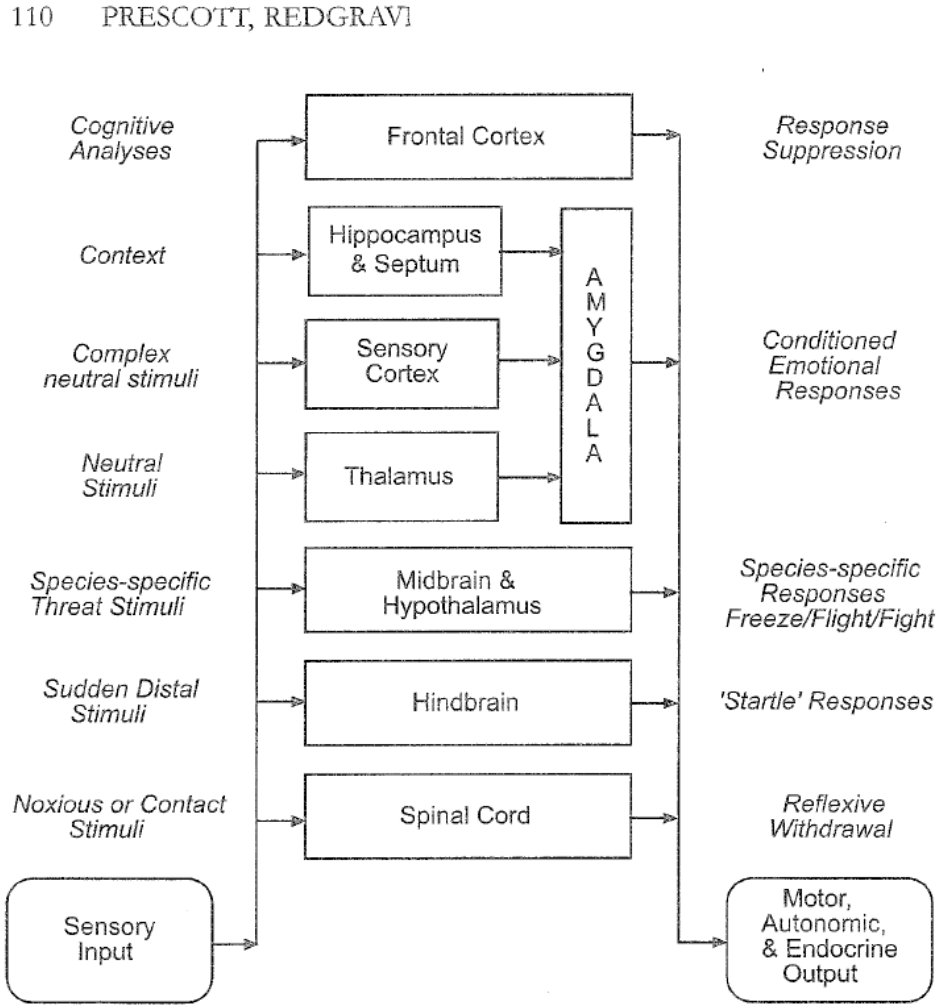




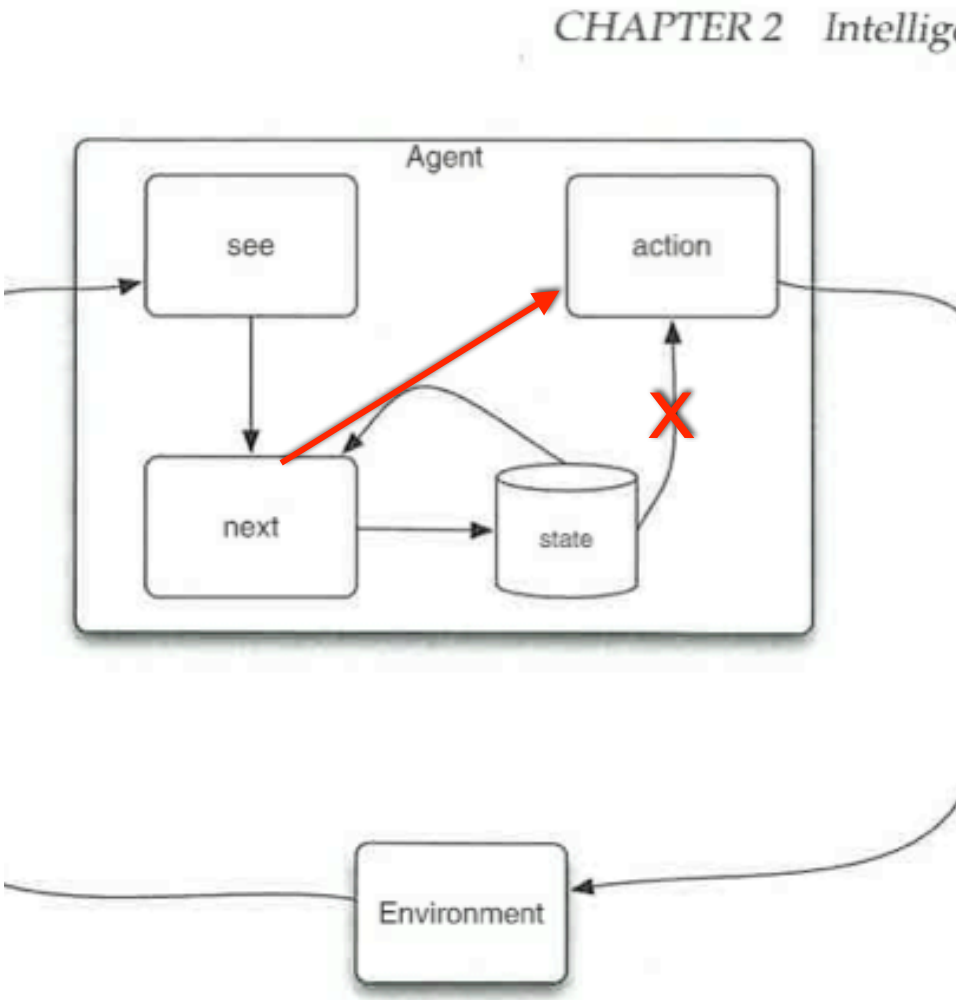
It’s possible to combine an internal sensorimotor loop with mechanisms for broadcast and competition, and thereby marry the *simulation hypothesis* with *global workspace theory*

Remember / Revision

Prescott after Brooks



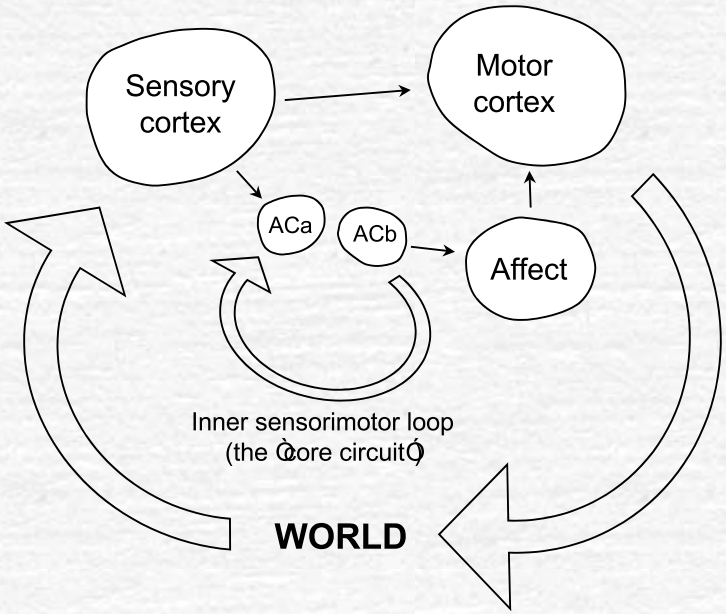
corrected Wooldridge

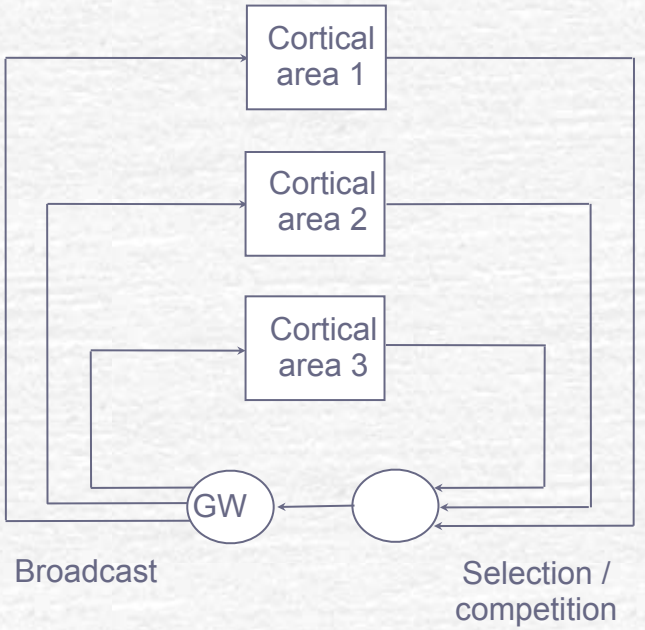


Science & Evolution

- Selection requires variation – occurs between existing options (and their combinations & mutations).
- History matters – understanding it helps explain what we think.
  - Some combination of what works well and what we were lucky someone thought of – culture.

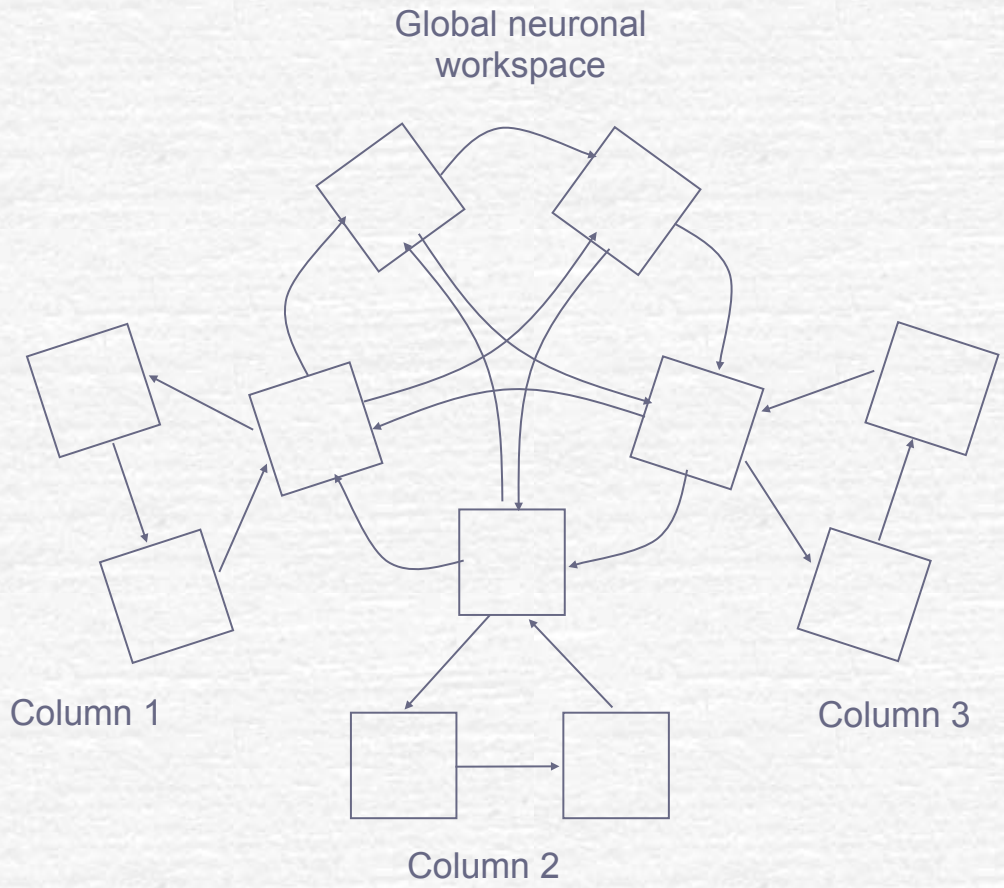
Combining a GW with Internal Simulation





It’s possible to combine an internal sensorimotor loop with mechanisms for broadcast and competition, and thereby marry the *simulation hypothesis* with *global workspace theory*

A Biologically Non-implausible Implementation



- Built out of spiking neurons with transmission delays
- Cortical columns comprise 32 × 32 fully connected nets
- Workspace nodes comprise 16 × 16 topographically mapped regions
- Cortical columns trained to associate successively presented pairs of images using STDP

Controlling a Robot

- The inner sensorimotor loop can be embedded in a larger system and used to control a robot
- This results in a form of “cognitively-enhanced” action selection

icing

- The implemented action selection architecture
  - Is based on salience and winner-takes-all
  - Imposes a veto at final motor output stage
  - Modulates salience as a result of internal simulation
  - Releases veto when salience exceeds a threshold

Pretty much Maes nets again.

- The parallelism of the GW architecture enables the inner loop to explore alternatives

Search vs Time

- Combinatorics is the problem, search is the only solution.
- The task of intelligence is to focus search.
  - Called bias (learning) or constraint (planning).
  - Most behaviour has no or little real-time search.
- For natural intelligence, most focus evolves.
  - Physical/cognitive constraints limit search space.

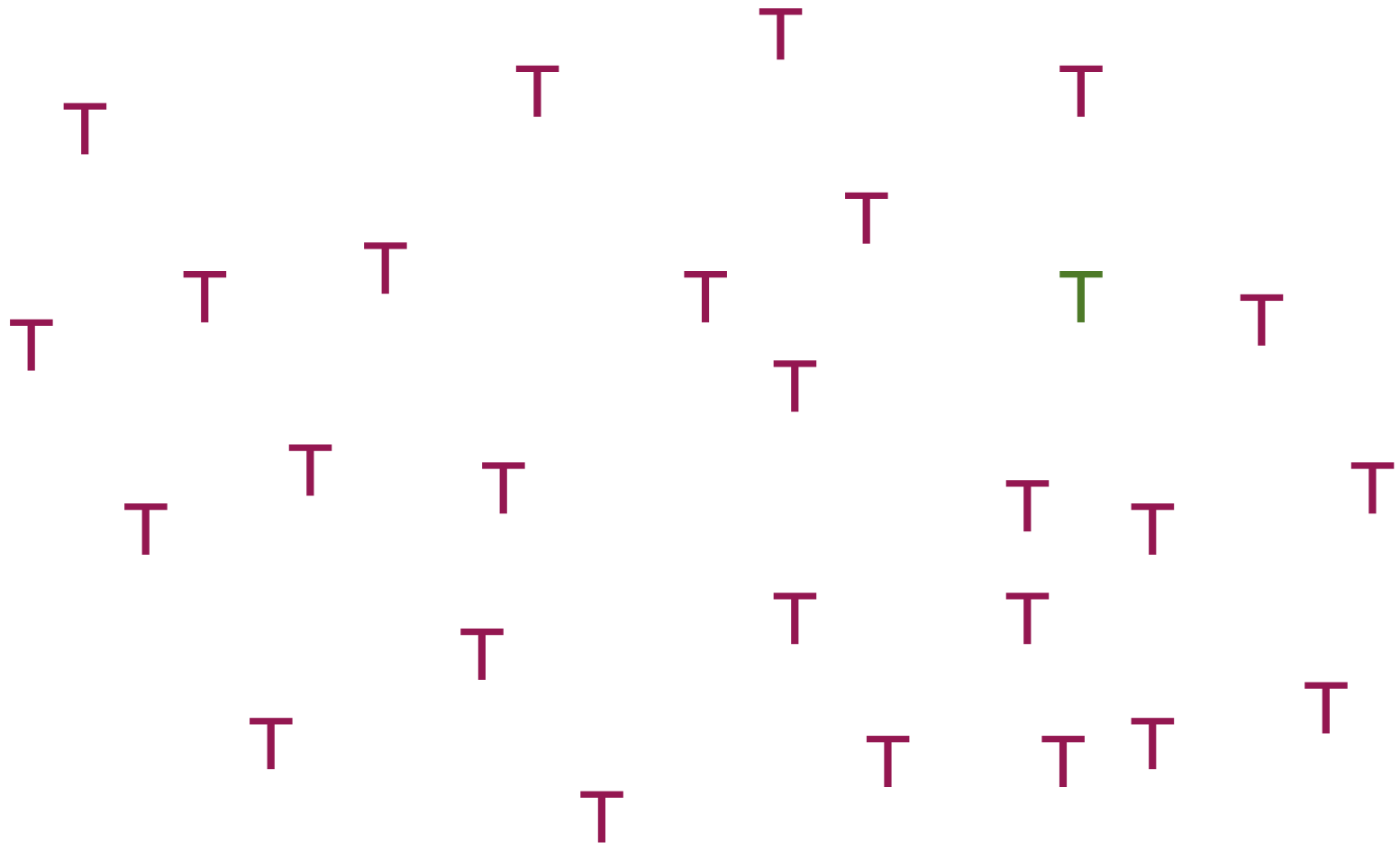
Hypothesis

Consciousness & cognition are that mental stuff that takes time.

(Treisman & Gelade, *Cognitive Psychology* 1980)

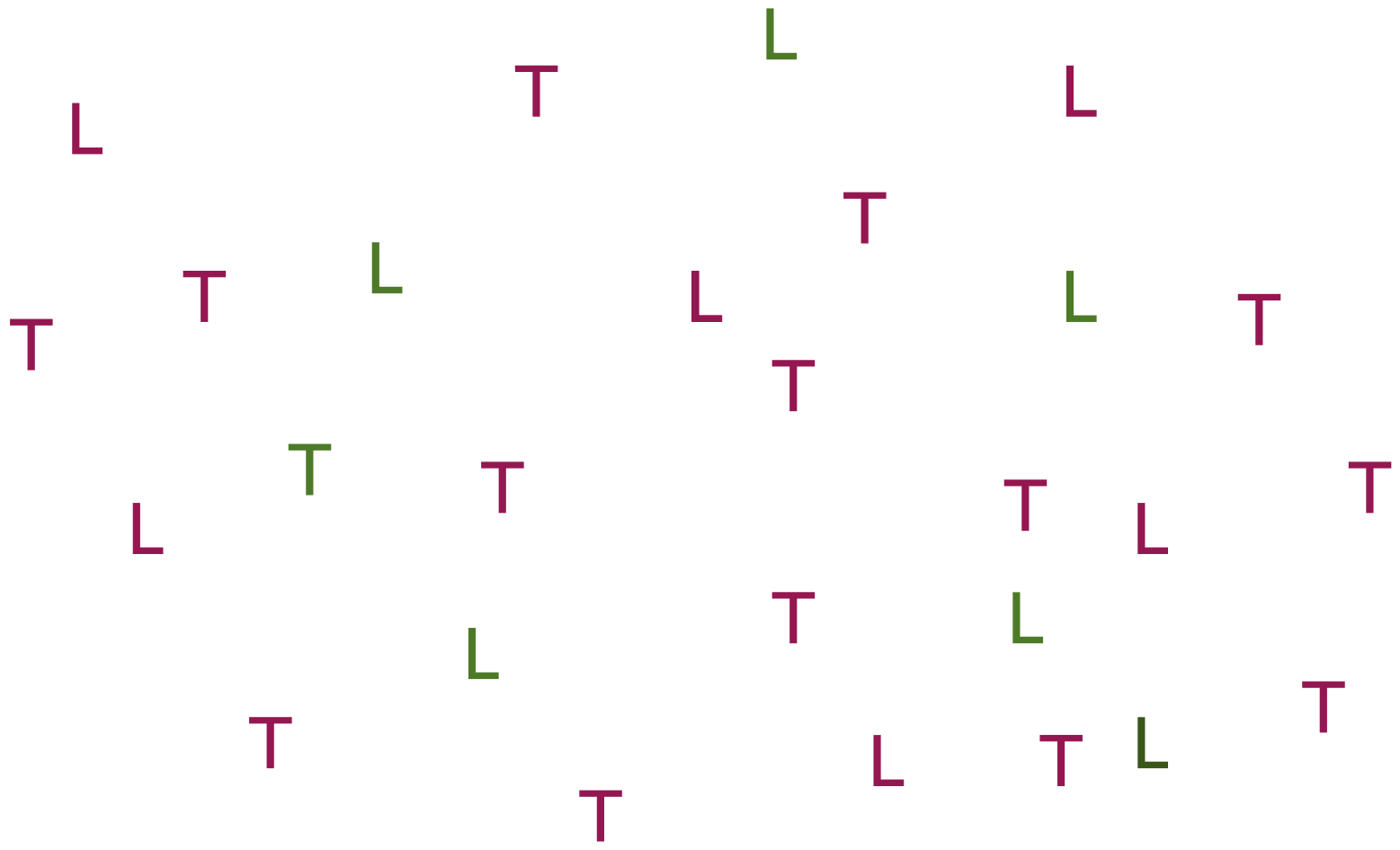
Ex 1: find the green T

Ex 1: find the green T



Ex 1: find the green T

Ex 1: find the green T



Time & Consciousness

- Sometimes time is determined by the number of steps you need to do (e.g. counting to yourself, searching a screen.)
- But sometimes it seems to be determined by something else…

Learning and Time

- Looking-time experiments rely on reaction-time delay being indicative of surprise.
- Flattening of reaction times correlates with failure to notice shift in reward schedule, but no impact on performance (Rapp *et al* 1998).

Looking time research e.g. Santos, Spelke

Allocating Time & Attention

I. Individuals allocate more time when less certain (Bryson 2009; 2010).

II. Species allocate in response to niche e.g. tamarins & insects (Hauser 1999).

III. Species allocate inversely with age (Rapp et al 1998, Bryson 2009; 2010).

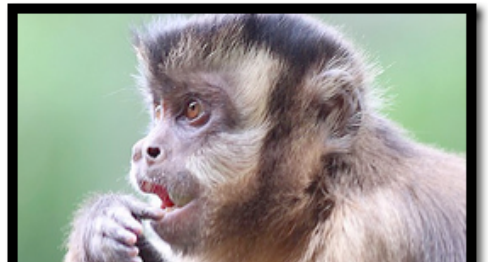
IV. Individuals allocate inversely with urgency (Shadlen and Newsome, 1998; Bogacz et al., 2006).

A Theory of Conscious Attention

- The basic function of conscious awareness is to update important models (learn).
  - Time is allocated in proportion to uncertainty by inhibiting action.
- Not to choose immediate action!
  - If new action is favoured due to model updates, may affect immediate behaviour.

Consciousness for AI

- Only need C if system learns, and learning relies on a bottlenecked cognitive resource.
- In this case, allocating C to tasks you are doing in proportion to how uncertain you are about them is a pretty good guess.
  - Also attend to other novel / unpredicted by your internal model events (deer in the headlights).



Point of Intervention

1. Action selection as usual.
2. Inhibit action expression while selected action is in mind, update models.
3. If new action becomes more salient, insight.
    - Flush plan & start over.
4. Update of models may not have immediate impact on behaviour.

Conclusions

Bryson 2011, 2012

- The basic function of awareness is not to choose actions, but to inhibit actions once selected and learn about their situation.
- A costly (in terms of time) allocation of resources for learning, varies in application by species and by individual situation.
- Easy to build.

Are There Already Conscious AI Systems?

Andrea Thomaz, MIT

Charlie Kemp, GA Tech

"If the best the roboticists can hope for is the creation of some crude, cheesy, second-rate artificial consciousness, they still win."

D. C. Dennett (1994), "The Practical Requirements for Making a Conscious Robot", *Philosophical Transactions: Physical Sciences and Engineering*, **349** p. 137 (133-146).

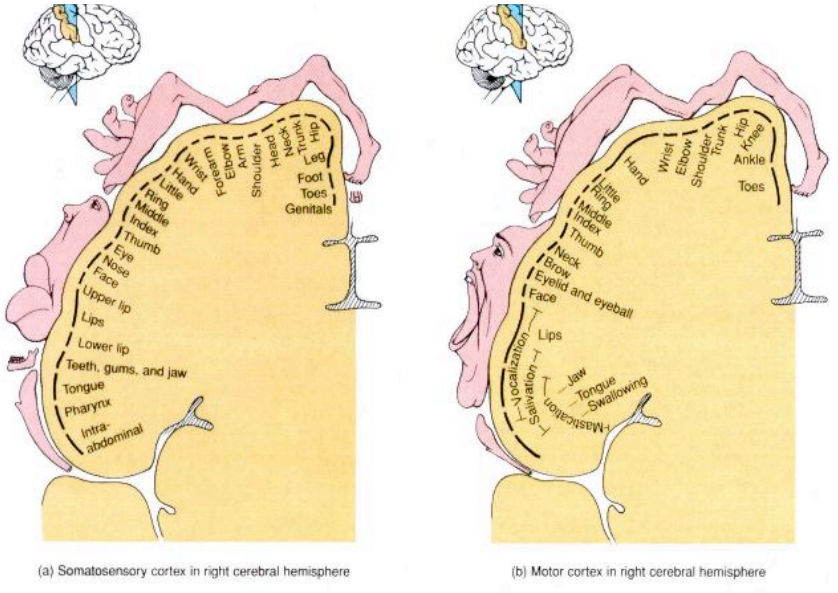
How does this relate to other theories of consciousness?

Roadmap for Conscious Machines

1. (-1) Disembodied
1. (0) Isolated
1. Decontrolled
2. Reactive
3. Adaptive
4. Attentional
5. Executive
6. Emotional
7. Self-conscious
8. Empathic
9. Social
10. Human-like
11. Super Conscious

Arrabales et al 2009

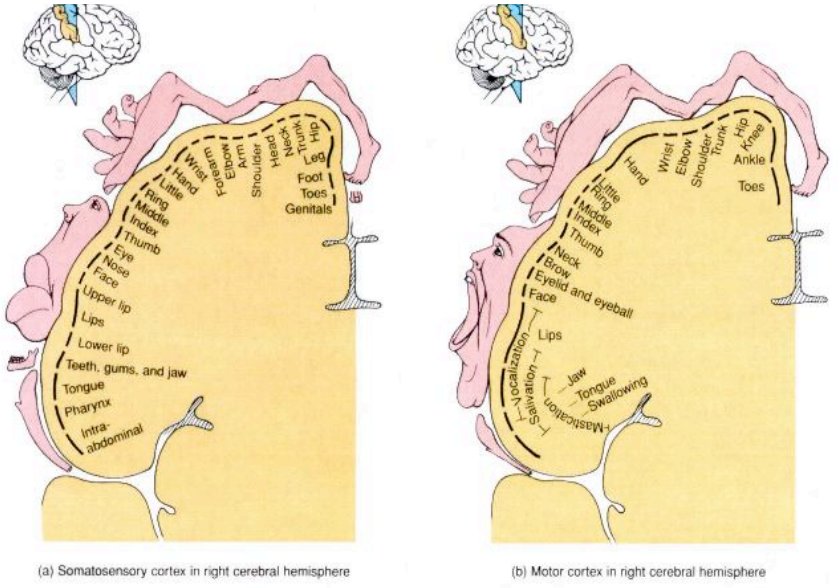
Self Consciousness



(a) Somatosensory cortex in right cerebral hemisphere (b) Motor cortex in right cerebral hemisphere

- Consciousness of self: limited like all consciousness to likely useful search space.
- Much facilitated in humans by language & instruction ⇒ probably less in other species.
- Google Search treats its own pages like other's: self-awareness neither necessary nor sufficient for consciousness.

Self Consciousness



(a) Somatosensory cortex in right cerebral hemisphere

(b) Motor cortex in right cerebral hemisphere

- Consciousness of self: limited like all consciousness to likely useful search space.
- Much facilitated in humans by language & instruction ⇒ probably less in other species.
- Google Search treats its own pages like other's: self-awareness neither necessary nor sufficient for consciousness.

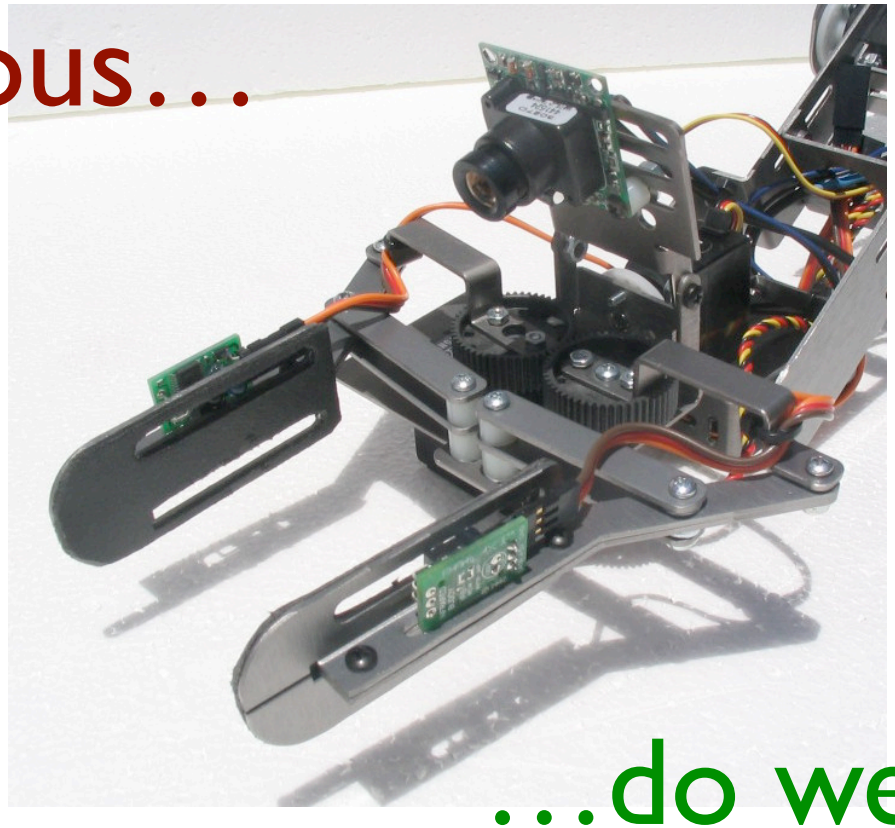
Language Helps

- Symbolic representation allows more compact and / or less emotionally-salient representations.
- Learn concepts from others; shared consciousness of events (Dennett 2008).
- Not a prerequisite for this basic functional component of action selection.

Ethics

- Consciousness: culturally-evolved concept of uncertain age and origin (Dennett, 2001).
  - May refer to no single psychological phenomenon.
- Ethics: Co-evolve with social order.
  - Much relies on assigning responsibility: covaries with but not determined by consciousness.

If robots are conscious…



…do we have to think harder about morality?

Gratuitous XKCD (Munroe 2012)

ICCS Conclusions

- Learned about autonomous intelligence by programming robots.
- Learned about interacting social intelligence (a little) by programming ABM.
- Learned a marketable skill by programming a game.
- Please teach me by filling in the unit review form – we really do read the free text!