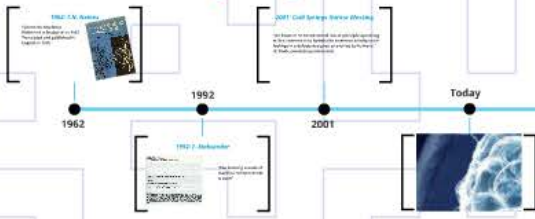
Machine Consciousness
Antonio Chella

Motivations

- Artifacts like us: consciousness, emotion and affect, experience, imagination, creativity (Robotics)
- Studying natural systems with computer laboratory models (Cognitive Science)
- Proficient machines (Intelligent Control)

A brief history



What is consciousness?

Consciousness consists of inner, qualitative, subjective states and processes of sentience or awareness. Consciousness, so defined, begins when we wake in the morning from a dreamless sleep and continues until we fall asleep again, die, go into a coma, or otherwise become "unconscious."

Axioms for Robot Consciousness

- I feel that I am a part of, but separate from an "out there" world
- I feel that my perception of the world mingles with feelings of past experience
- My experience of the world is selective and purposeful
- I am thinking ahead all the time in trying to decide what to do next
- I have feelings, emotions and moods that determine what I do

Information Integration

Homework

Models for MC

- Consciousness as Information Integration
- Consciousness as Introspection/Monitoring
- Consciousness as Internal Model

Introspection/monitor models

- Hierarchy of modules in the computational minds
- Low level modules related with reactive input-outputs
- High level modules related with deliberative planning, reasoning, ...
- Monitor modules

Internal Models

- An intelligent agent has an internal model of itself and of the external world
- Capability to simulate the external environment and the body actions
- Generation of expectations
- "Small scale model" of external reality (Craik)
- Popperian creatures (Dennett)

UNIVERSITÀ DEGLI STUDI DI PALERMO

What is consciousness?

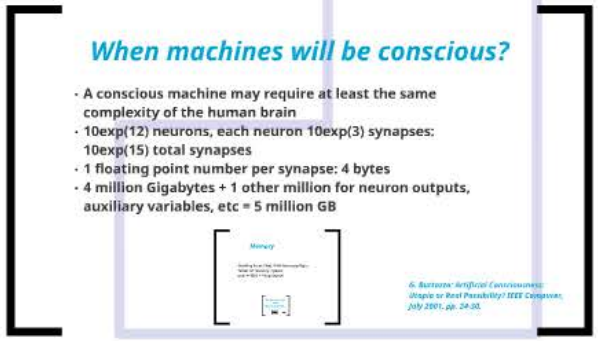
Consciousness consists of inner, qualitative, subjective states and processes of sentience or awareness. Consciousness, so defined, begins when we wake in the morning from a dreamless sleep and continues until we fall asleep again, die, go into a coma, or otherwise become "unconscious."

J.R. Searle: Consciousness, Annu. Rev. Neurosci. 2000. 23:557–578

Motivations

- Artifacts like us: consciousness, emotion and affect, experience, imagination, creativity (Robotics)
- Studying natural systems with computer laboratory models (Cognitive Science)
- Proficient machines (Intelligent Control)

R. Sanz, I. Lopez & J. Bermejo-Alonso: A Rationale and Vision for Machine Consciousness in Complex Controllers, in: A. Chella, R. Manzotti (eds.), Artificial Consciousness, Imprint Academic, 2007.



The New Experimental Science of Physical Cognitive Systems

AI, Robotics, Neuroscience and Cognitive Sciences under a New Name with the Old Philosophical Problems?

Fabio Bonsignorio

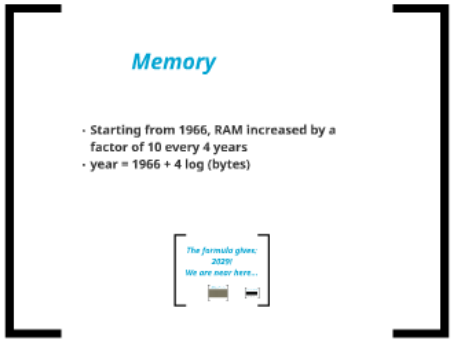
Abstract. It is likely that in AI, Robotics, Neuroscience and Cognitive Sciences, what we need is an integrated approach putting together concepts and methods from fields so far considered well distinct like non linear dynamics, information, computation and control theory as well as general AI, psychology, cognitive sciences in general, neurosciences and system biology. These disciplines usually share many problems, but have very different languages and experimental methodologies. It is thought that while tackling with many serious 'hard core' scientific issues it is imperative, probably a necessary (pre) requisite, that we do serious efforts to clarify and merge the underlying paradigms, the proper methodologies, the metrics and success criteria of this new branch of science. Many of these questions have already been approached by philosophy, but they acquire in this context a scientific nature: e.g.: Is it possible cognition without consciousness? And without 'sentience'? In the context of AI and neuroscience research various definition of consciousness have been proposed (for example by Tononi, [44], to quote an example liked by the author). How they relate to the previous and contemporary philosophical analysis? Sometimes scientists may look as poor philosophers, and the opposite: philosophers may look as poor scientists, yet, the critical passages of history of science during a paradigm change or the birth of a new discipline have often involved a highly critical conceptual analysis intertwined with scientific and mathematical advancements. The scientific enterprise is now somehow close to unbundle the basic foundation of our consciousness and of our apperception of reality, and, it is clear that there are some circularity issues with the possible 'explanations', at least.

Fabio Bonsignorio
University Carlos III of Madrid and Heron Robots

V.C. Müller (Ed.): Philosophy and Theory of Artificial Intelligence, SAPERE 5, pp. 133–150.
springerlink.com

When machines will be conscious?

- **A conscious machine may require at least the same complexity of the human brain**
- **10exp(12) neurons, each neuron 10exp(3) synapses: 10exp(15) total synapses**
- **1 floating point number per synapse: 4 bytes**
- **4 million Gigabytes + 1 other million for neuron outputs, auxiliary variables, etc = 5 million GB**



G. Buttazzo: Artificial Consciousness: Utopia or Real Possibility? IEEE Computer, July 2001, pp. 24-30.

Memory

- Starting from 1966, RAM increased by a factor of 10 every 4 years
- year = 1966 + 4 log (bytes)

The formula gives:
2029!
We are near here...

The formula gives: 2029!

We are near here...

In 2008! (before iPhone era)

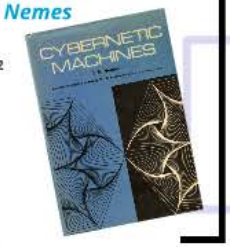
Do we really want conscious robots?

In 2008! (before iPhone era)

A brief history

1962: T.N. Nemes

Cybernetic Machines
Published in Budapest in 1962
Translated and published in English in 1970



2001: Cold Springs Harbor Meeting

"we know of no fundamental law or principle operating in this universe that forbids the existence of subjective feelings in artefacts designed or evolved by humans."
(C. Koch, concluding comments)

1962

1992

2001

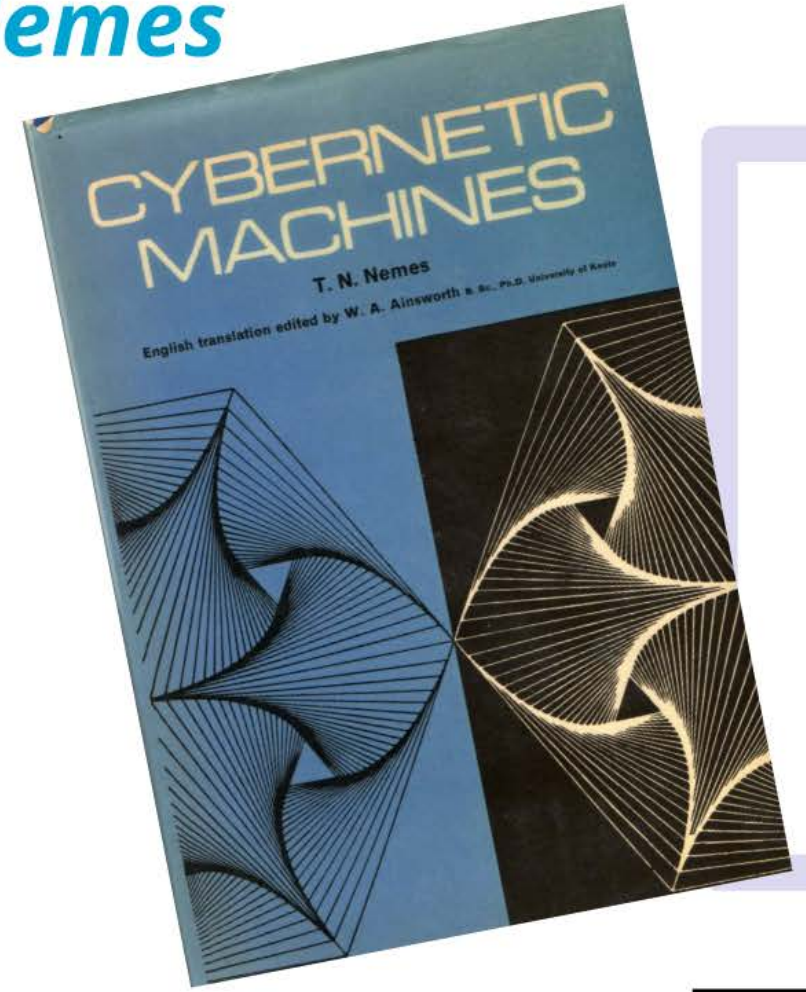
Today

1992: I. Aleksander

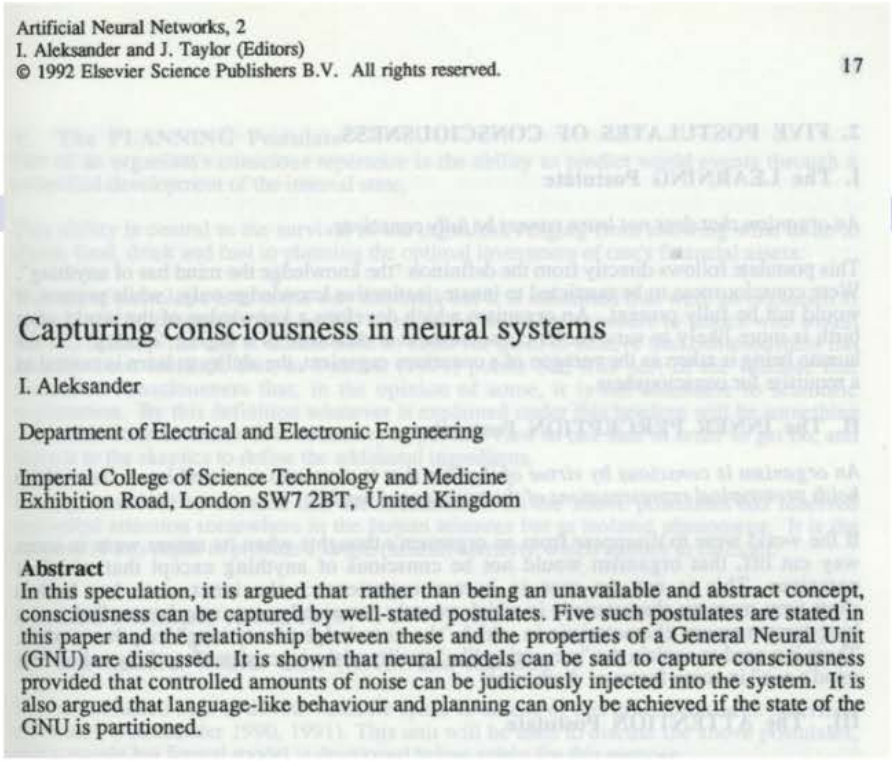
"The hunting season of machine consciousness is open"

1962: T.N. Nemes

Cybernetic Machines
Published in Budapest in 1962
Translated and published in English in 1970



1992: I. Aleksander



Artificial Neural Networks, 2
I. Aleksander and J. Taylor (Editors)

Capturing consciousness in neural systems

I. Aleksander

Department of Electrical and Electronic Engineering

Imperial College of Science Technology and Medicine
Exhibition Road, London SW7 2BT, United Kingdom

Abstract
In this speculation, it is argued that rather than being an unavailable and abstract concept, consciousness can be captured by well-stated postulates. Five such postulates are stated in this paper and the relationship between these and the properties of a General Neural Unit (GNU) are discussed. It is shown that neural models can be said to capture consciousness provided that controlled amounts of noise can be judiciously injected into the system. It is also argued that language-like behaviour and planning can only be achieved if the state of the GNU is partitioned.

"The hunting season of machine consciousness is open"

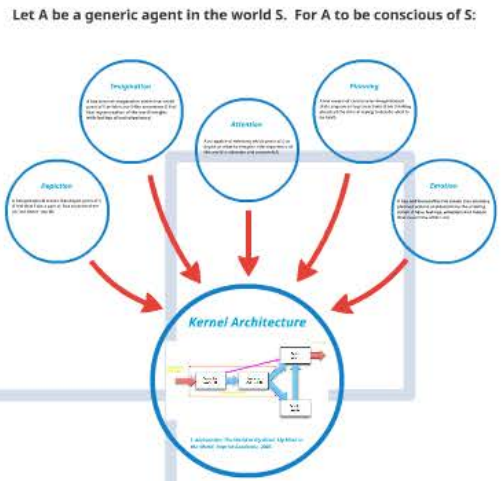
2001: Cold Springs Harbor Meeting

“we know of no fundamental law or principle operating in this universe that forbids the existence of subjective feelings in artefacts designed or evolved by humans.”
(C. Koch, concluding comments)


ieee spectrum 06.08
RTIFICIAL
NSCIOUSNESS
RE-ENGINEERING EARTH
Who decides when to hack the planet?
REALLY SMART CARS
No driver required
DAMP SIDE OF THE MOON
It's wetter than we thought
NewScientist
International Journal of
MACHINE CONSCIOUSNESS
Volume 1 · Number 1 · June 2009
Editor-in-Chief
Antonio Chella
World Scientific
ARTIFICIAL INTELLI
MEDICINE
machine
consciousness
edited by
owen holland

Axioms for Robot Consciousness

- I feel that I am a part of, but separate from an "out there" world
- I feel that my perception of the world mingles with feelings of past experience
- My experience of the world is selective and purposeful
- I am thinking ahead all the time in trying to decide what to do next
- I have feelings, emotions and moods that determine what I do



I. Aleksander & B. Dunmall: Axioms and Tests for the Presence of Minimal Consciousness in Agents, Journal of Consciousness Studies, 10, 4-5, pp.7-18, 2003

Let A be a generic agent in the world S. For A to be conscious of S:

Imagination

A has internal imagination states that recall parts of S or fabricate S-like sensations (I feel that my perception of the world mingles with feelings of past experience)

Attention

A is capable of selecting which parts of S to depict or what to imagine. (My experience of the world is selective and purposeful)

Planning

A has means of control over imaginational state sequences to plan actions (I am thinking ahead all the time in trying to decide what to do next)

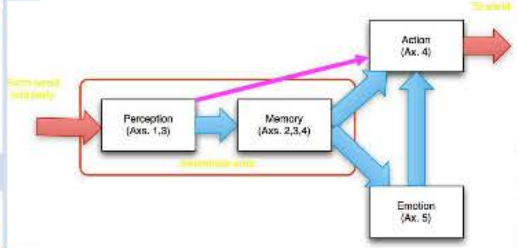
Depiction

A has perceptual states that depict parts of S (I feel that I am a part of, but separate from an "out there" world)

Emotion

A has additional affective states that evaluate planned actions and determine the ensuing action (I have feelings, emotions and moods that determine what I do)

Kernel Architecture



I. Aleksander: The World in My Mind, My Mind in the World, Imprint Academic, 2005

Depiction

A has perceptual states that depict parts of S (I feel that I am a part of, but separate from an "out there" world)

Imagination

A has internal imagination states that recall parts of S or fabricate S-like sensations (I feel that my perception of the world mingles with feelings of past experience)

Attention

A is capable of selecting which parts of S to depict or what to imagine. (My experience of the world is selective and purposeful)

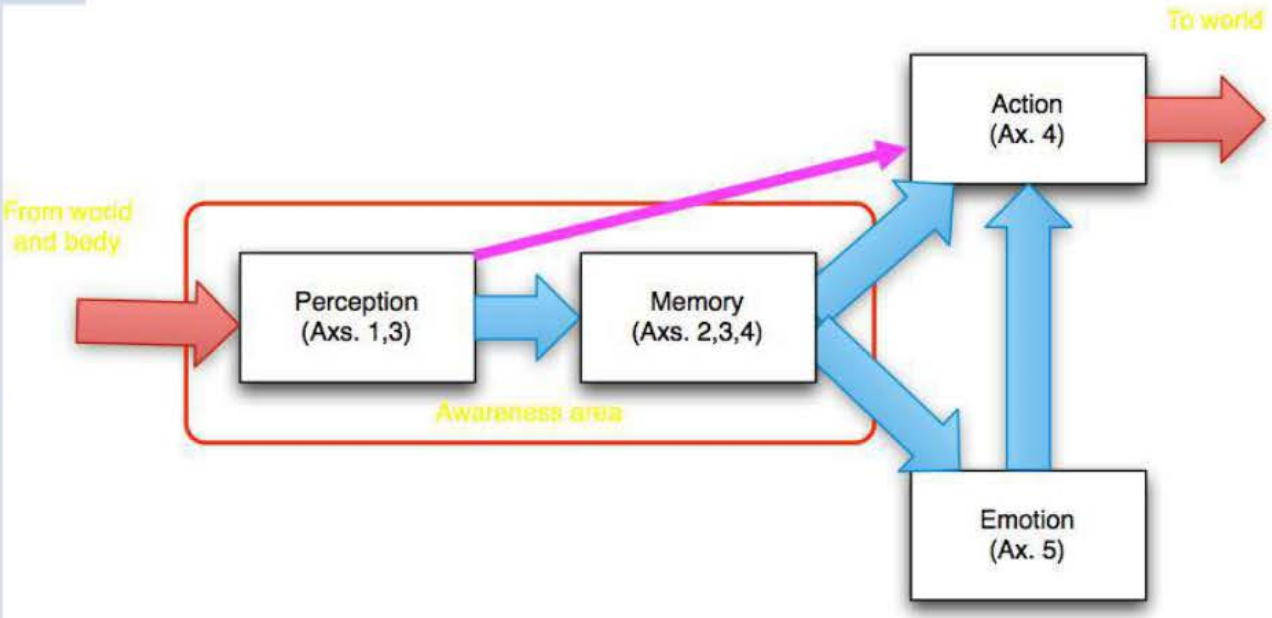
Planning

A has means of control over imaginational state sequences to plan actions (I am thinking ahead all the time in trying to decide what to do next)

Emotion

A has additional affective states that evaluate planned actions and determine the ensuing action (I have feelings, emotions and moods that determine what I do)

Kernel Architecture



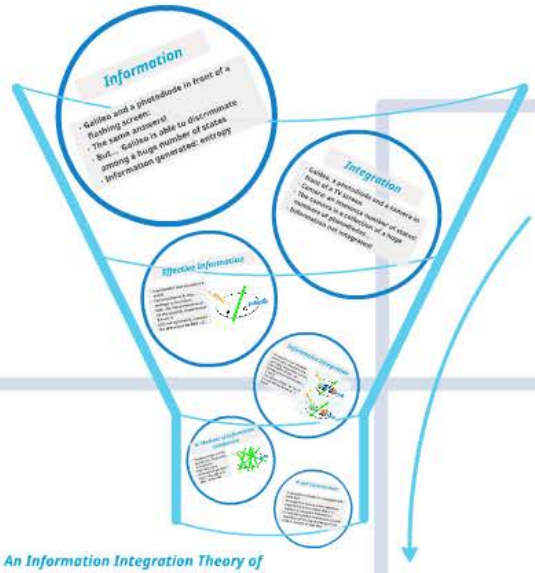
I. Aleksander: The World in My Mind, My Mind in the World, Imprint Academic, 2005

Models for MC

- **Consciousness as Information Integration**
- **Consciousness as Introspection/Monitoring**
- **Consciousness as Internal Model**

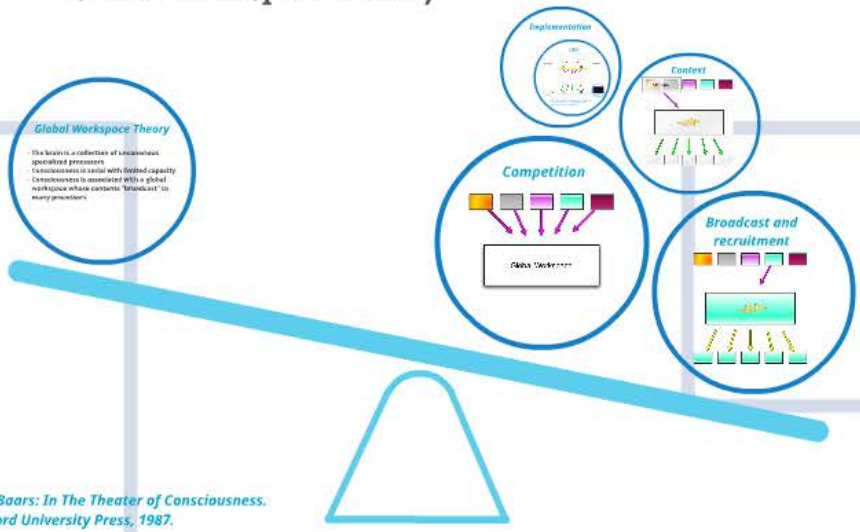
Information Integration

Tononi



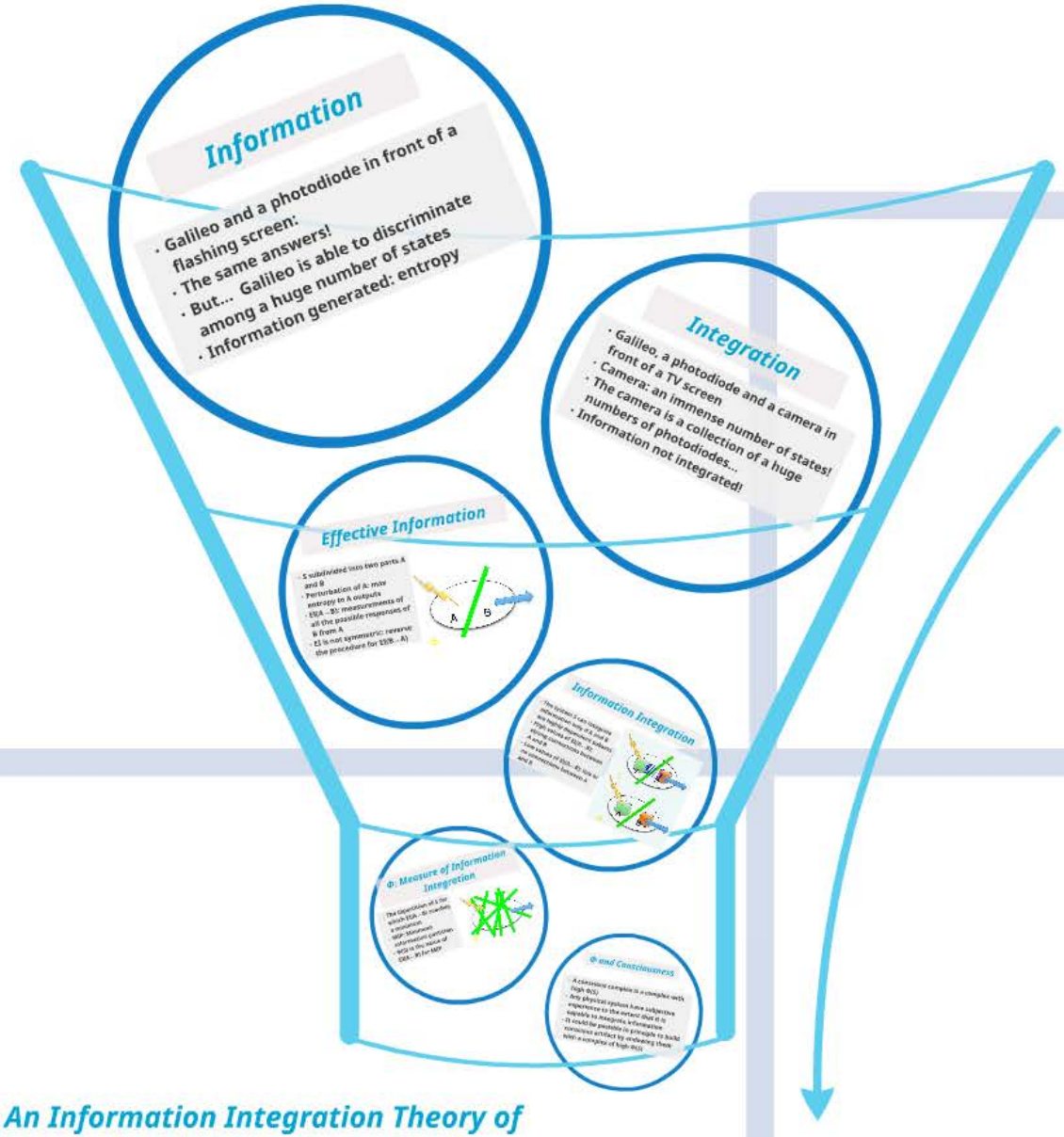
G. Tononi: An Information Integration Theory of Consciousness, BMC Neuroscience, 5-42, 2004

Global Workspace Theory



B.J. Baars: In The Theater of Consciousness. Oxford University Press, 1987.

Tononi



G. Tononi: An Information Integration Theory of Consciousness, BMC Neuroscience, 5-42, 2004

B.J
Ox

Information

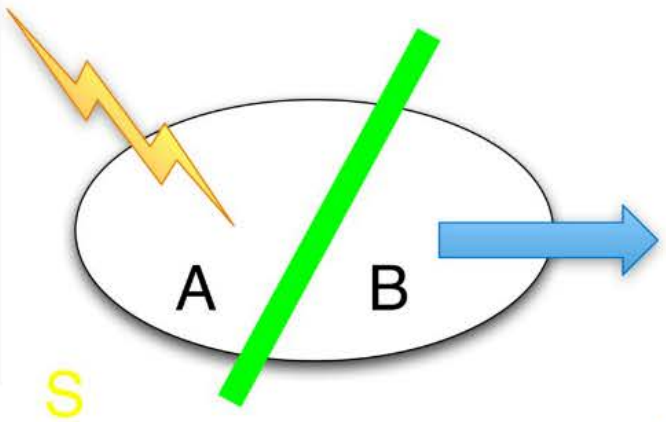
- Galileo and a photodiode in front of a flashing screen:
- The same answers!
- But… Galileo is able to discriminate among a huge number of states
- Information generated: entropy

Integration

- **Galileo, a photodiode and a camera in front of a TV screen**
- **Camera: an immense number of states!**
- **The camera is a collection of a huge numbers of photodiodes...**
- **Information not integrated!**

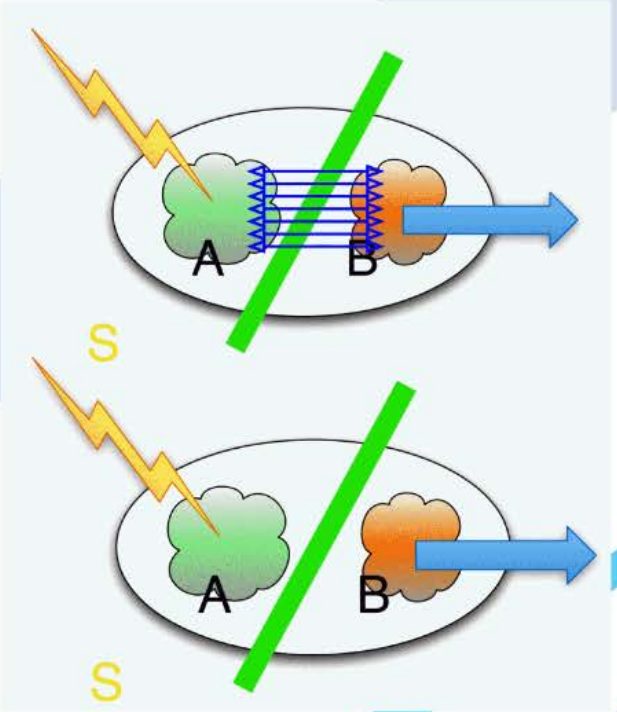
Effective Information

- S subdivided into two parts A and B
- Perturbation of A: max entropy to A outputs
- EI(A → B): measurements of all the possible responses of B from A
- EI is not symmetric: reverse the procedure for EI(B → A)



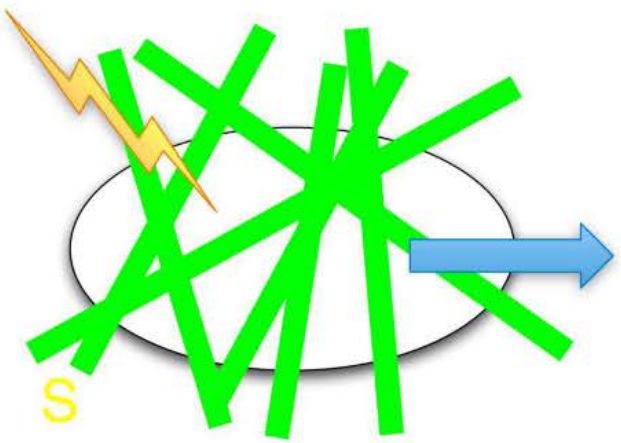
Information Integration

- The system S can integrate information only if A and B are highly dependent subsets
- High values of EI(A ↔ B): strong connections between A and B
- Low values of EI(A ↔ B): low or no connections between A and B



Φ: Measure of Information Integration

- The bipartition of S for which EI(A ↔ B) reaches a minimum
- MIP: Minimum information partition
- Φ(S) is the value of EI(A ↔ B) for MIP

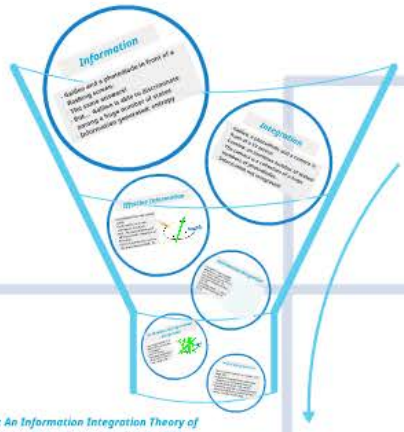


Φ and Consciousness

- A conscious complex is a complex with high Φ(S)
- Any physical system have subjective experience to the extent that it is capable to integrate information
- It could be possible in principle to build conscious artifact by endowing them with a complex of high Φ(S)

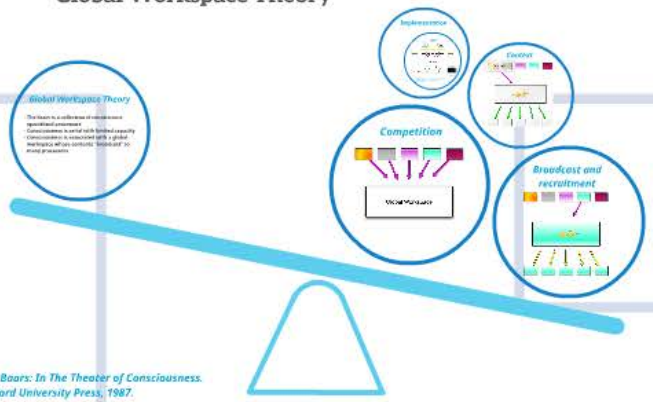
Information Integration

Tononi



G. Tononi: An Information Integration Theory of Consciousness, BMC Neuroscience, 5-42, 2004

Global Workspace Theory

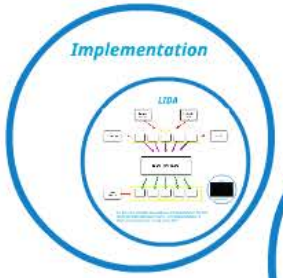


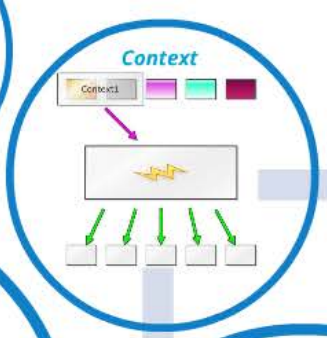
B.J. Baars: In The Theater of Consciousness. Oxford University Press, 1987.

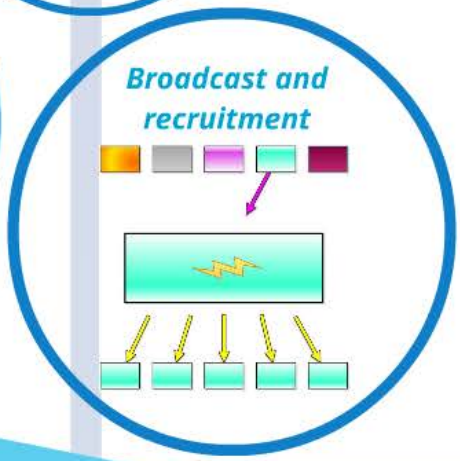
Global Workspace Theory

Global Workspace Theory

- The brain is a collection of unconscious specialized processors
- Consciousness is serial with limited capacity
- Consciousness is associated with a global workspace whose contents "broadcast" to many processors





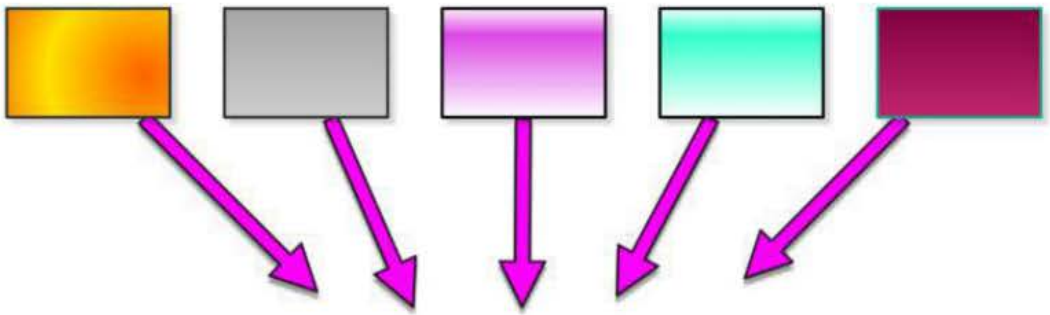


B.J. Baars: In The Theater of Consciousness. Oxford University Press, 1987.

Global Workspace Theory

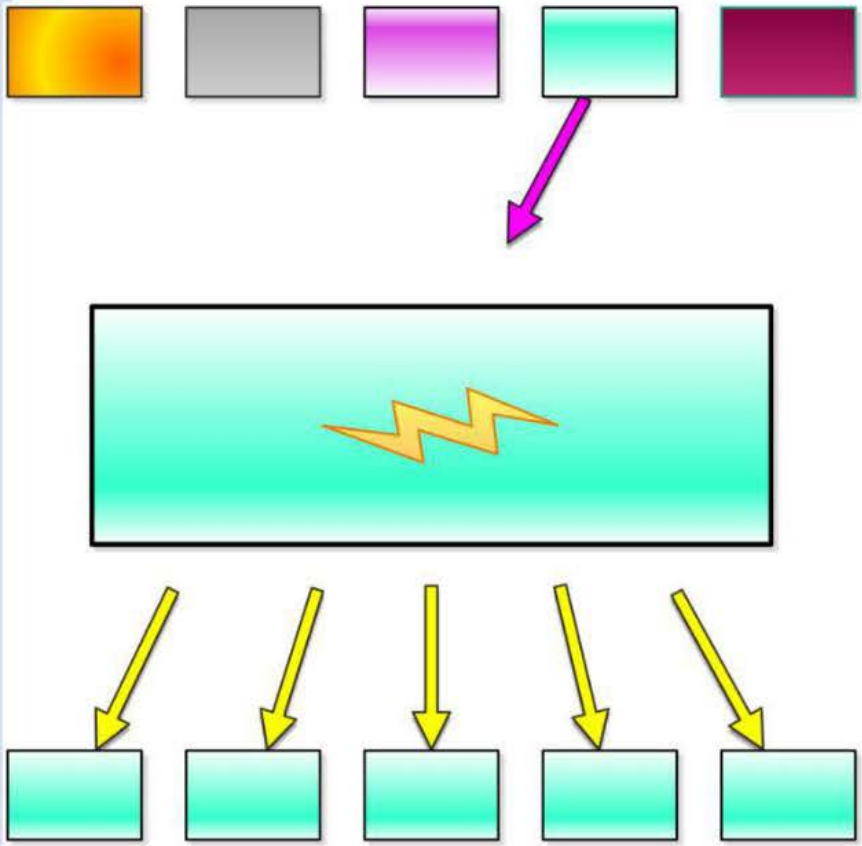
- **The brain is a collection of unconscious specialized processors**
- **Consciousness is serial with limited capacity**
- **Consciousness is associated with a global workspace whose contents "broadcast" to many processors**

Competition



Global Workspace

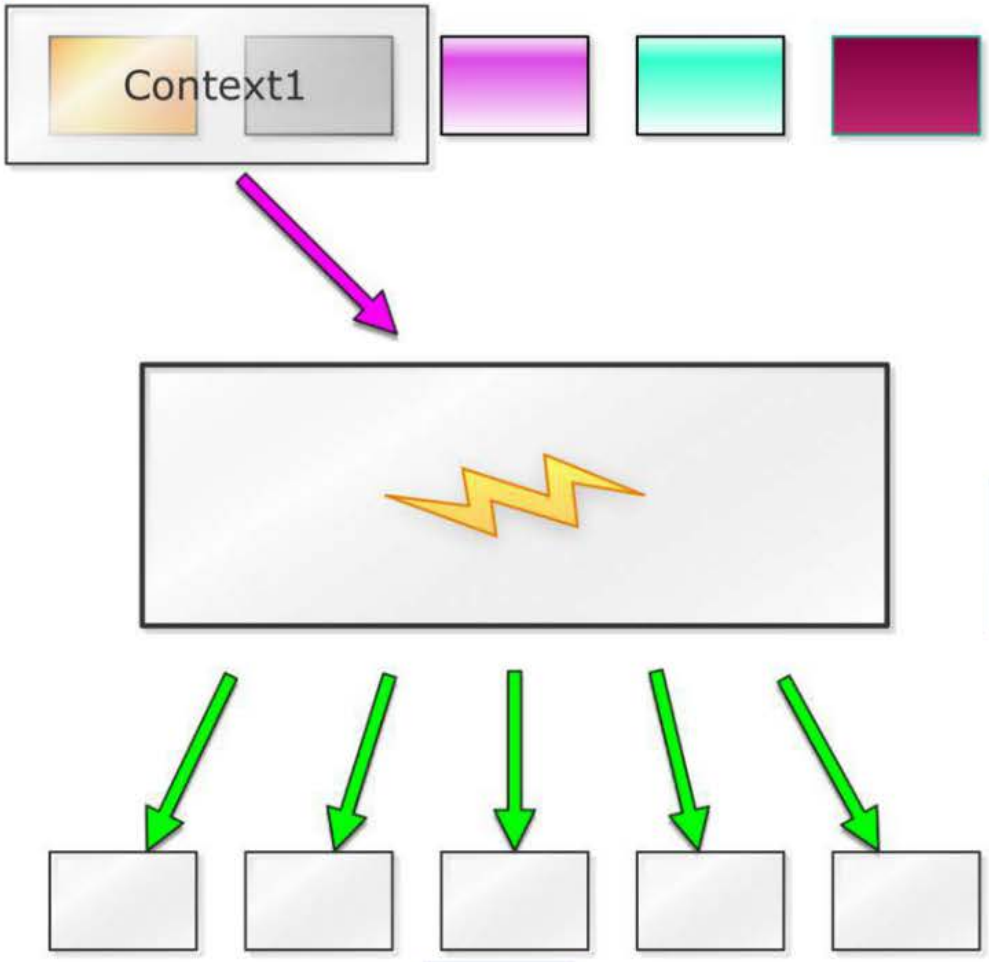
Broadcast and recruitment



Attention

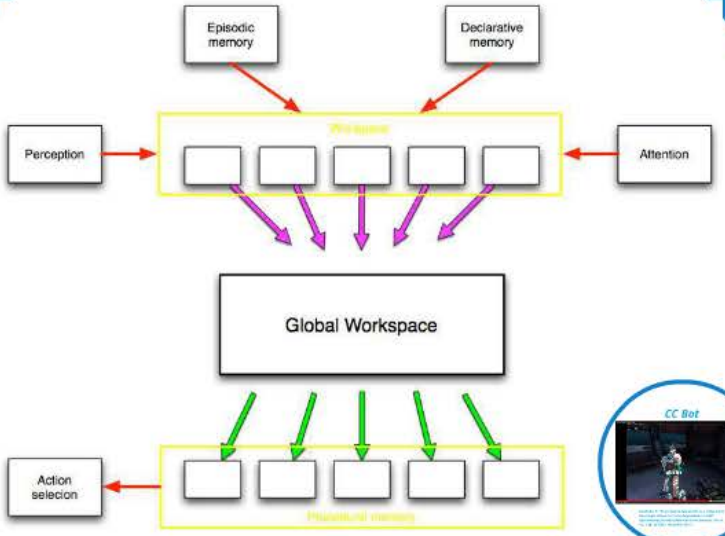
CC Bot

Context



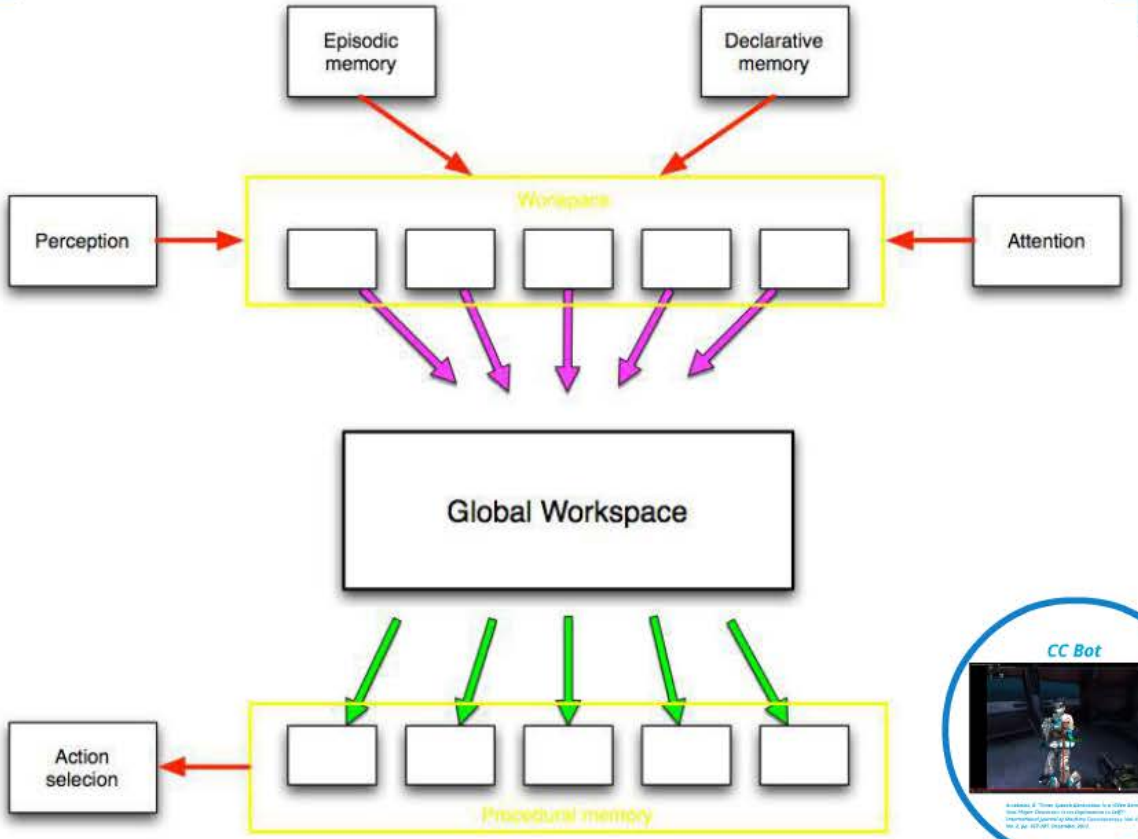
Implementation

LIDA



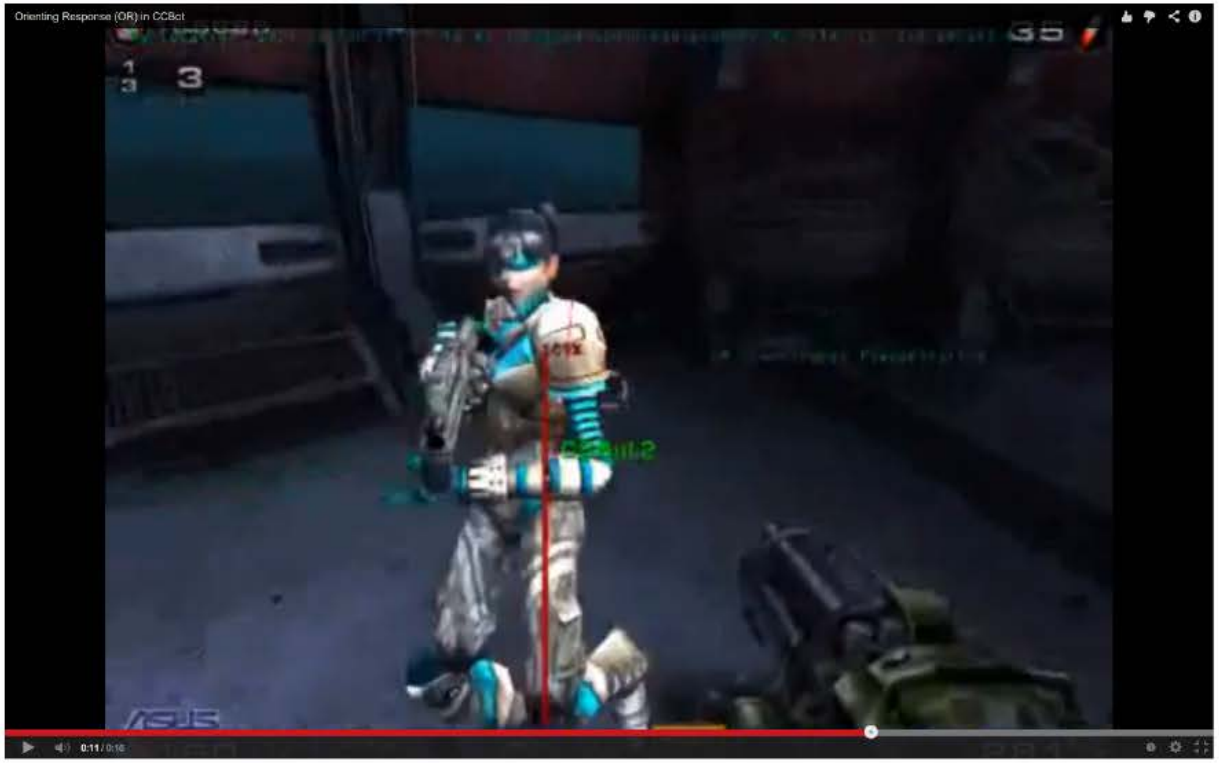
B.J. Baars & S. Franklin: Consciousness is Computational: The LIDA Model of Global Workspace Theory, International Journal of Machine Consciousness, 1, 1, pp. 23-32, 2009

LIDA



B.J. Baars & S. Franklin: Consciousness is Computational: The LIDA Model of Global Workspace Theory, International Journal of Machine Consciousness, 1, 1, pp. 23-32, 2009

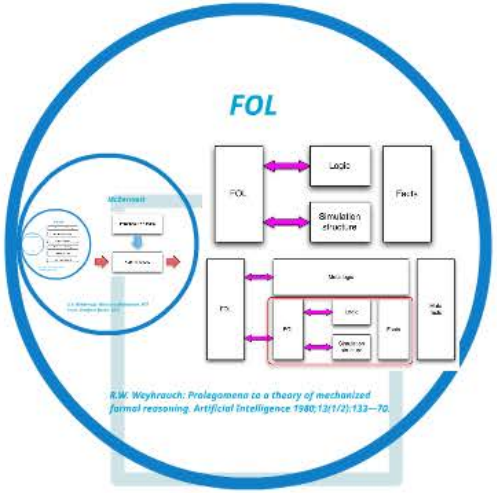
CC Bot

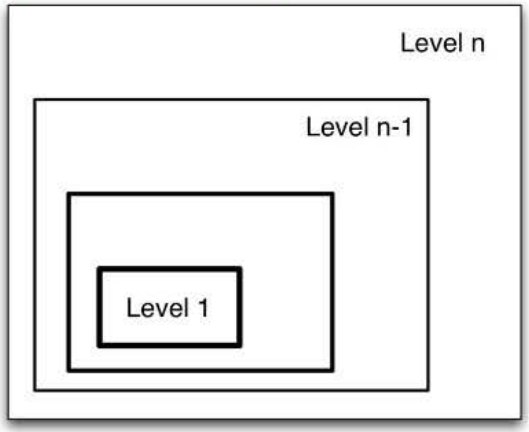


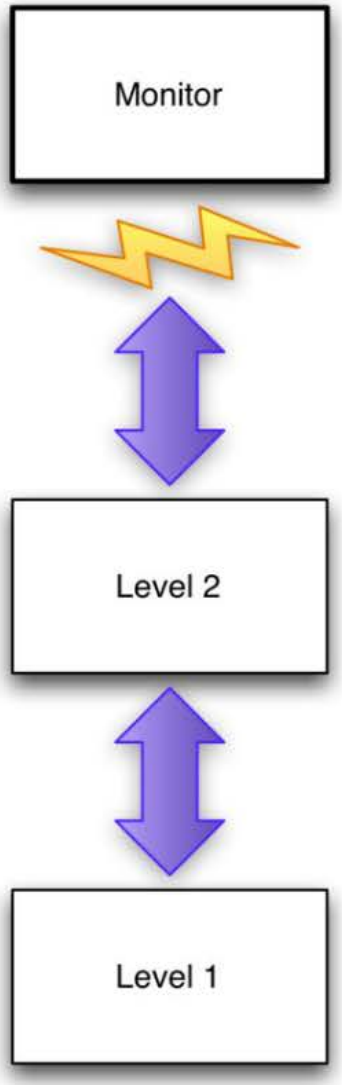
Arrabales, R. "Inner Speech Generation in a Video Game Non-Player Character: From Explanation to Self?". International Journal of Machine Consciousness. Vol. 4. No. 2. pp. 367-381. December 2012.

Introspection/monitor models

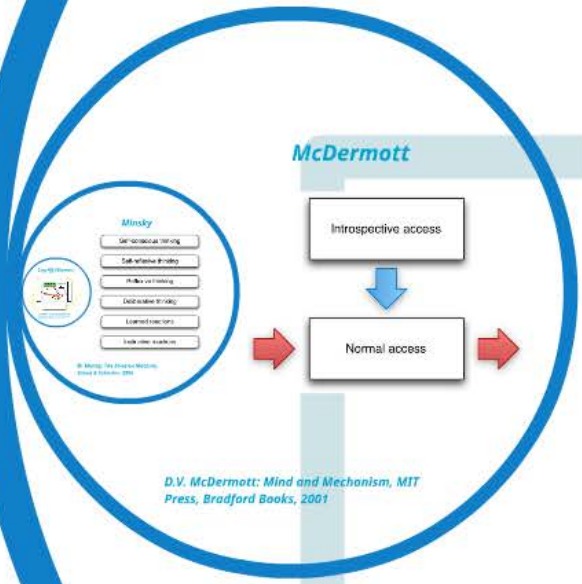
- Hierarchy of modules in the computational minds
- Low level modules related with reactive input-outputs
- High level modules related with deliberative planning, reasoning, ...
- Monitor modules

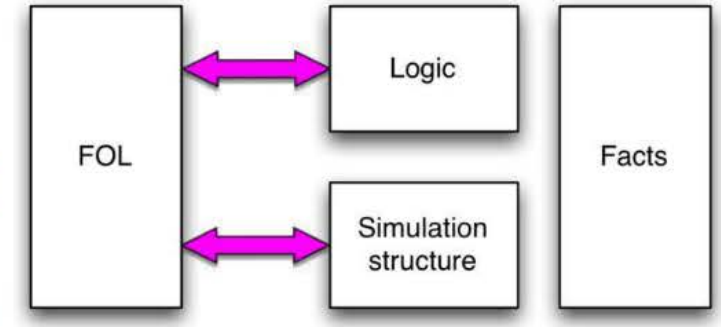


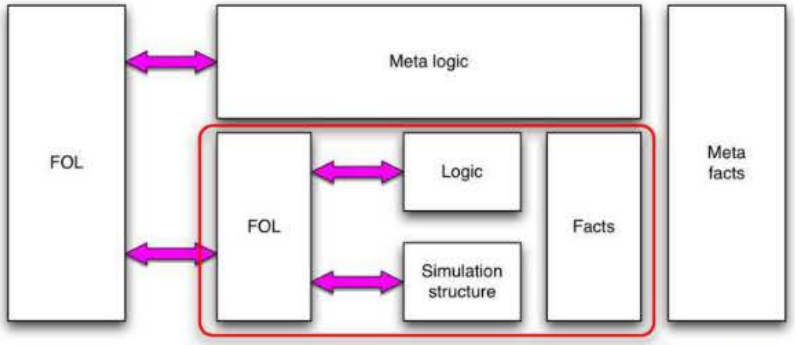




FOL







R.W. Weyhrauch: Prolegomena to a theory of mechanized formal reasoning. Artificial Intelligence 1980;13(1/2):133—70.

McDermott

Introspective access

Normal access

Minsky

Self-conscious thinking

Self-reflexive thinking

Reflexive thinking

Deliberative thinking

Learned reactions

Instinctive reactions

Cog-Aff (Sloman)

M. Minsky: The Emotion Machine, Simon & Schuster, 2006

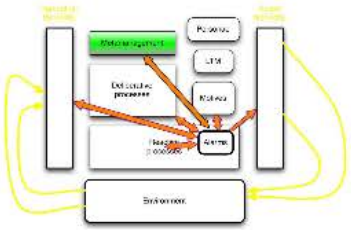
D.V. McDermott: Mind and Mechanism, MIT Press, Bradford Books, 2001

FOL

Minsky

- Self-conscious thinking
- Self-reflexive thinking
- Reflexive thinking
- Deliberative thinking
- Learned reactions
- Instinctive reactions

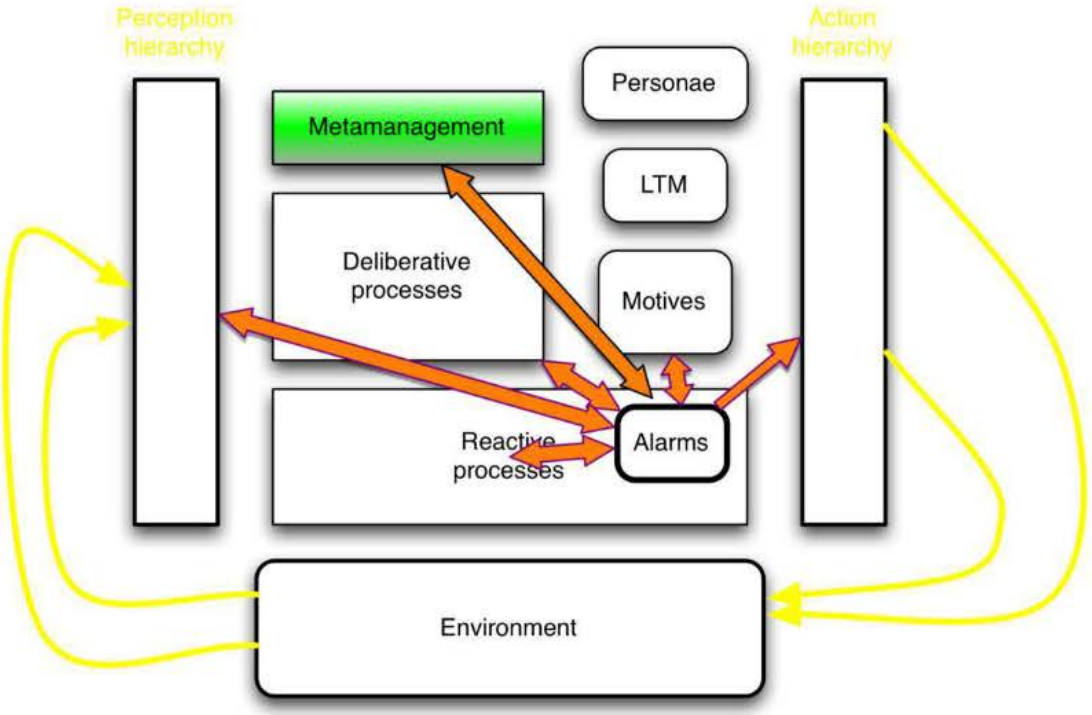
Cog-Aff (Sloman)



A. Sloman and Ron Chrisley: Virtual Machines and Consciousness, Journal of Consciousness Studies, 10, no. 4-5, pp. 133-172, 2003.

M. Minsky: The Emotion Machine, Simon & Schuster, 2006

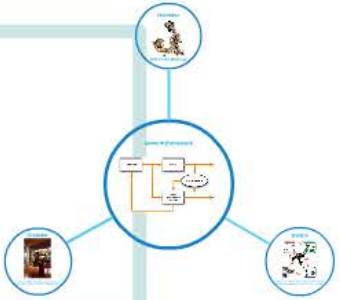
Cog-Aff (Sloman)

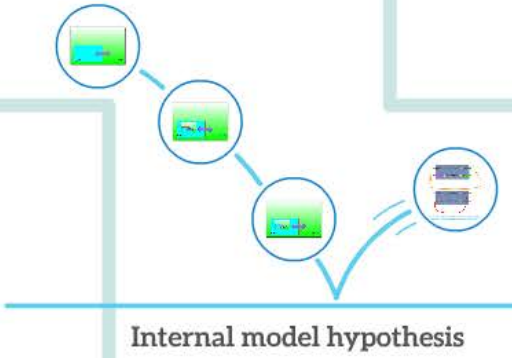


A. Sloman and Ron Chrisley: Virtual Machines and Consciousness, Journal of Consciousness Studies, 10, no. 4-5, pp. 133-172, 2003.

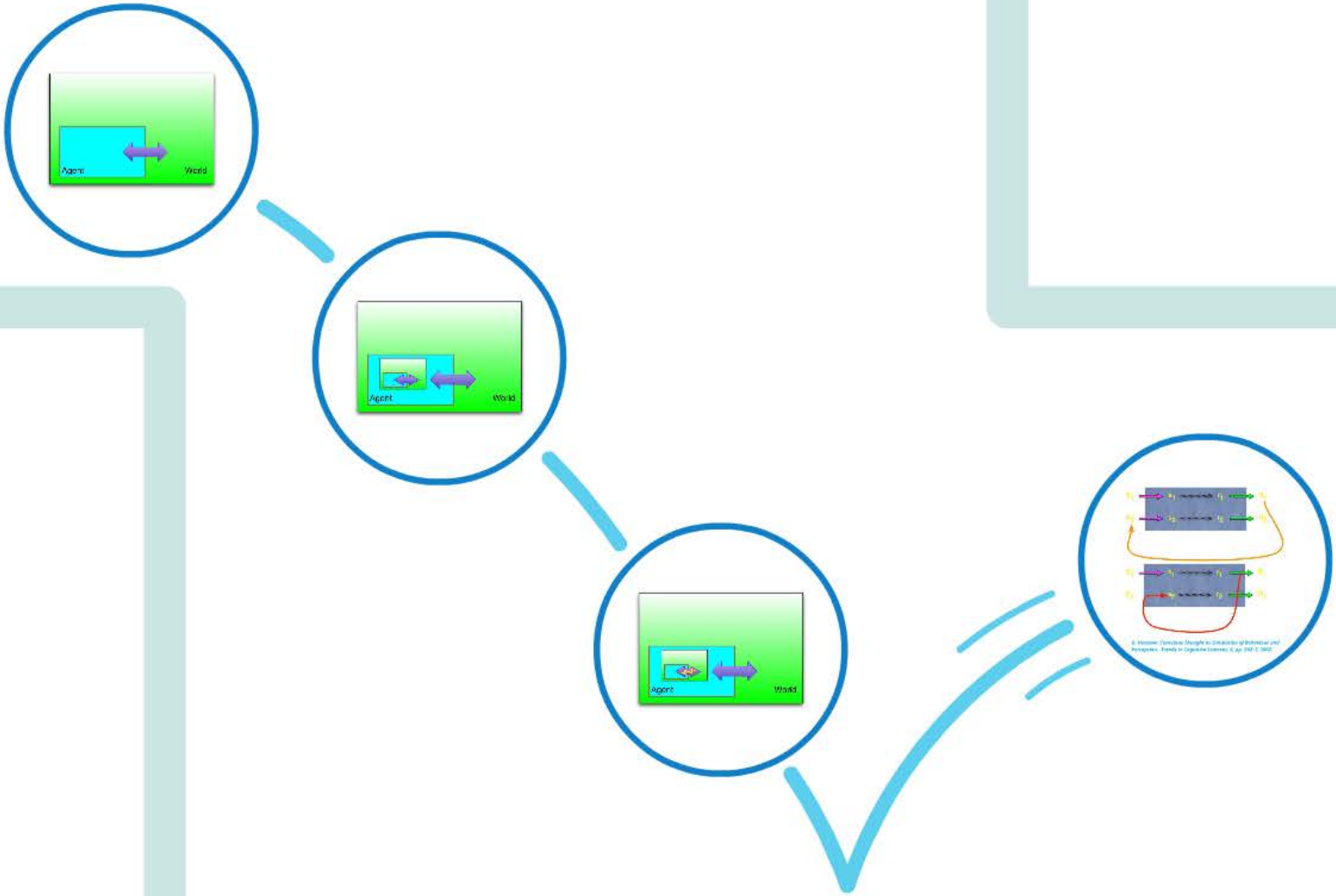
Internal Models

- An intelligent agent has an internal model of itself and of the external world
- Capability to simulate the external environment and the body actions
- Generation of expectations
- “Small scale model” of external reality (Craik)
- Popperian creatures (Dennett)



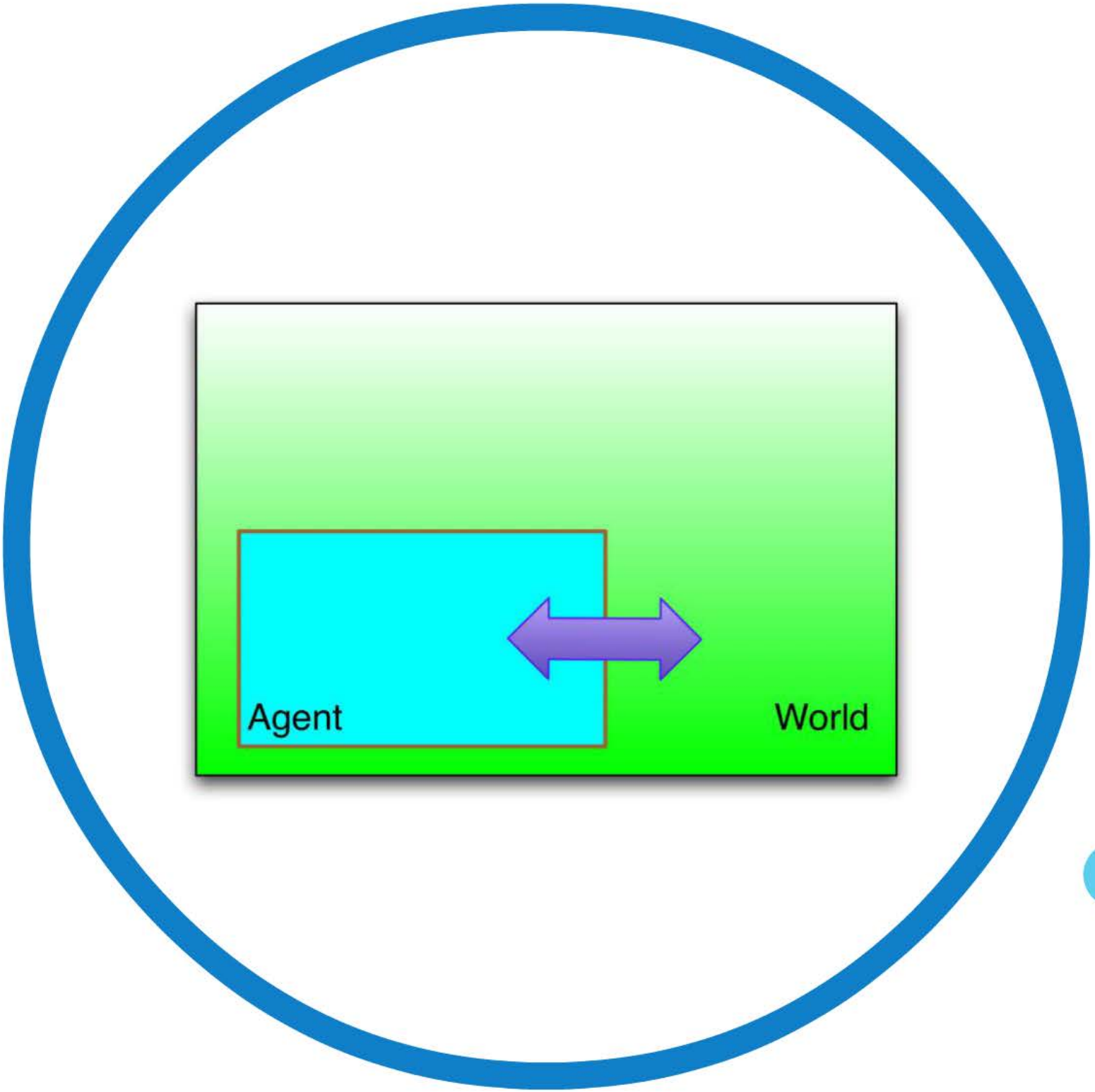


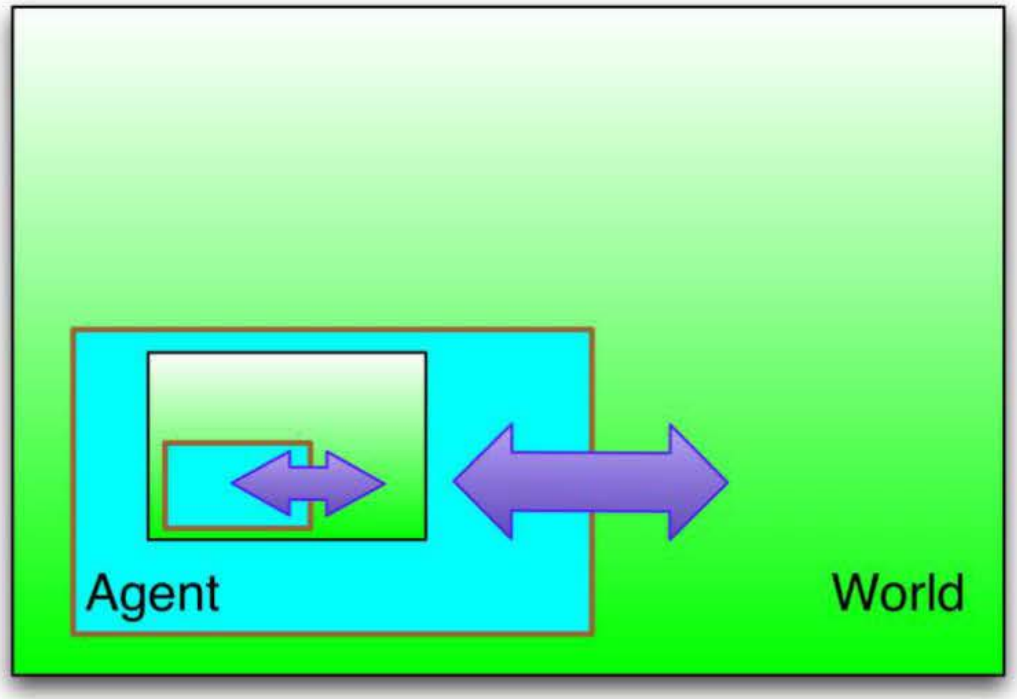
O. Holland, R. Knight, & R. Newcombe: A Robot-Based Approach to Machine Consciousness, in: A. Chella, R. Manzotti (eds.): Artificial Consciousness, Imprint Academic, 2007

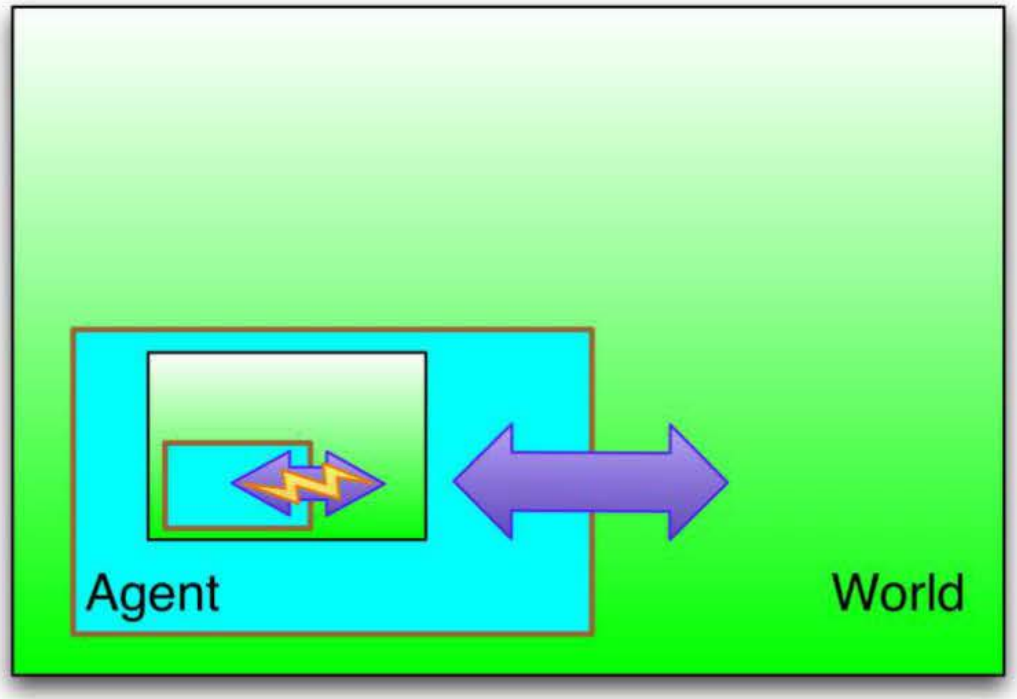


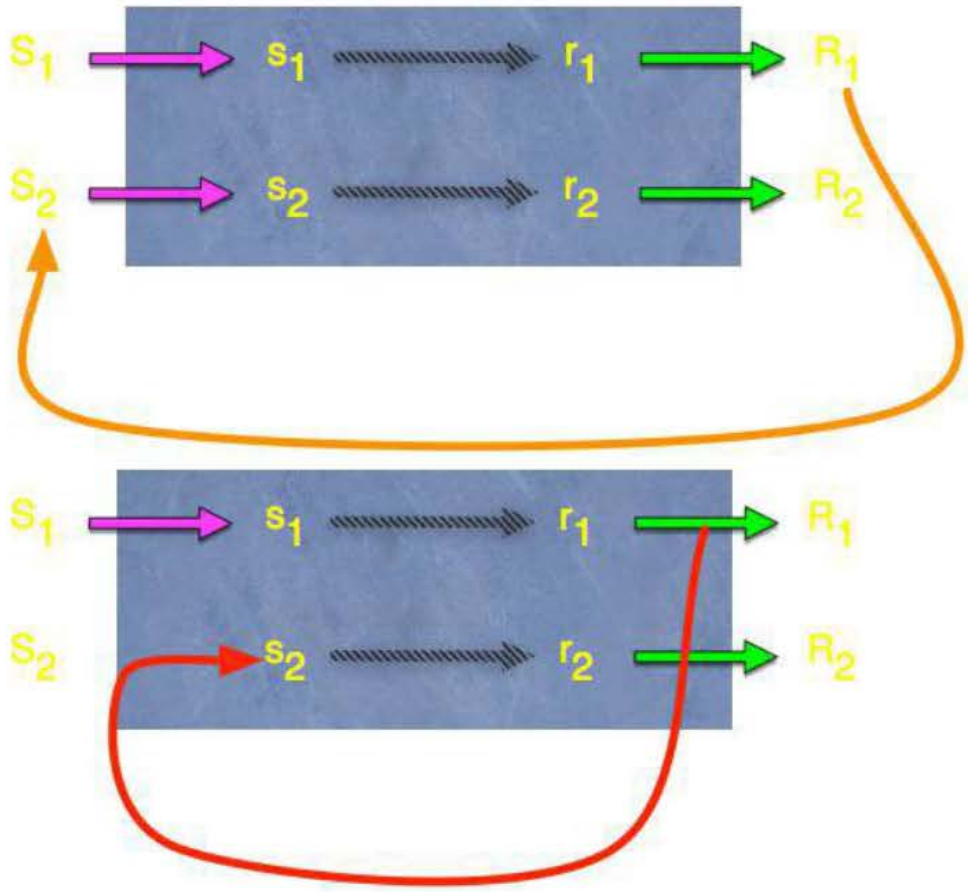
Internal model hypothesis

O. Holland, R. Knight, & R. Newcombe: A Robot-Based Approach to Machine Consciousness, in: A. Chella, R. Manzotti (eds.): Artificial Consciousness, Imprint Academic, 2007


Agent
World


Agent
World


Agent
World

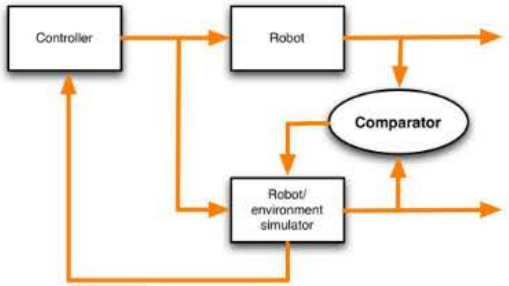


G. Hesslow: Conscious Thought as Simulation of Behaviour and Perception, Trends in Cognitive Sciences, 6, pp. 242–7, 2002.

EcceRobot

http://eccerobot.org/

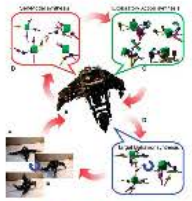
General framework



CiceRobot

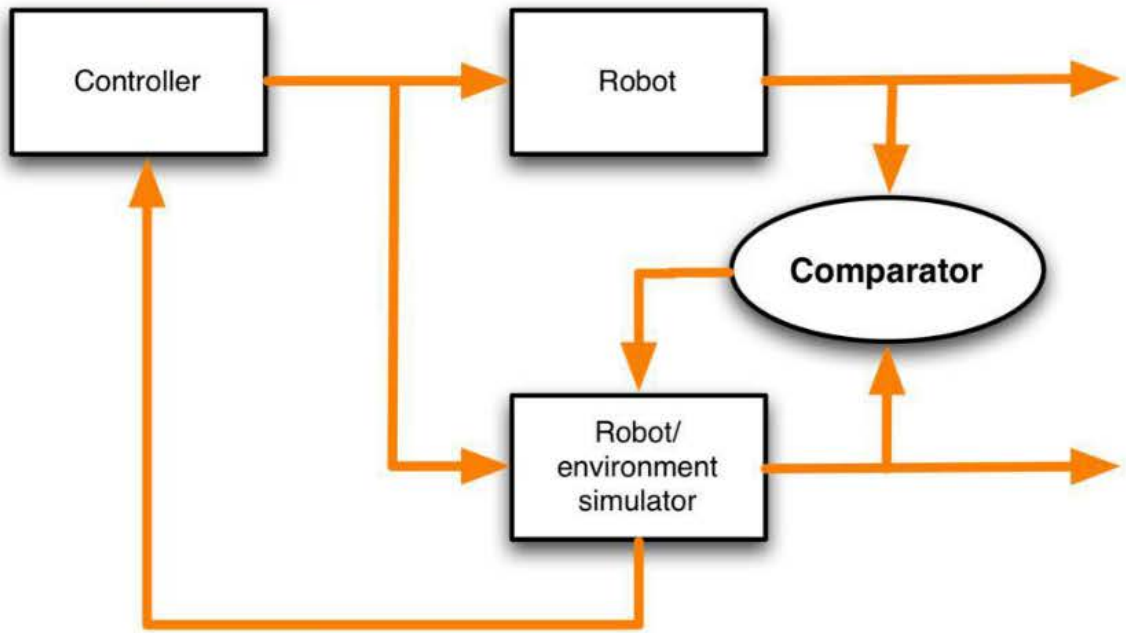
A. Chella, I. Macaluso: The perception loop in CiceRobot, a museum guide robot, Neurocomputing 72 760- 766, 2009

Starfish



J. Bongard, V. Zykov, H. Lipson: Resilient Machines Through Continuous Self-Modeling, Science, 314, 1118-1121, 2006

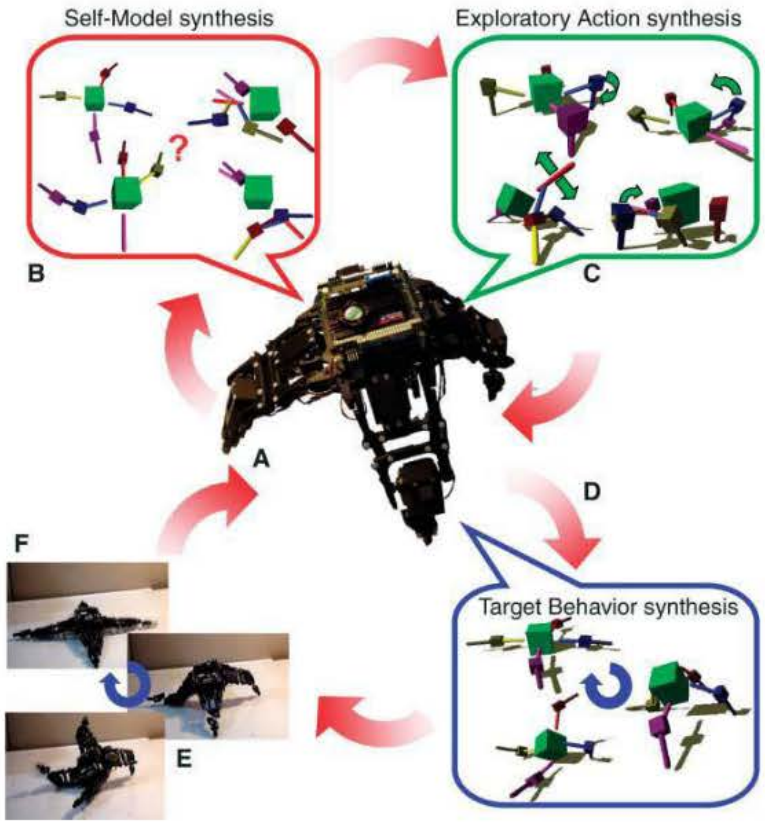
General framework



EcceRobot

http://eccerobot.org/

Starfish



J. Bongard, V. Zykov, H. Lipson: Resilient Machines Through Continuous Self-Modeling, Science, 314, 1118-1121, 2006

CiceRobot

A. Chella, I. Macaluso: The perception loop in CiceRobot, a museum guide robot, Neurocomputing 72 760–766, 2009

Homework

Intro

- D. Ledge Thinks..., Penguin, 2002
- S. Blackmore Consciousness, A Very Short Introduction, Oxford University Press, 2005

Machine Consciousness

- O. Holland (ed.) Machine Consciousness, Imprint Academic, 2003
- A. Chella, R. Manzotti (eds.) Artificial Consciousness, Imprint Academic, 2007
- Book Series on Machine Consciousness:
- http://www.worldscientific.com/series/smc

Websites

- http://www.consciousrobots.com/ (thanks to Raul Arrabales)
- Association for the Scientific Study of Consciousness:
- http://www.theassc.org/
- Center for Consciousness Studies - University of Arizona:
- http://www.consciousness.arizona.edu/

Journals

- International Journal of Machine Consciousness (currently discontinued)
- http://www.worldscinet.com/ijmc/
- Journal of Consciousness Studies
- http://www.imprint.co.uk/jcs.html
- Consciousness and Cognition
- www.elsevier.com/locate/concog

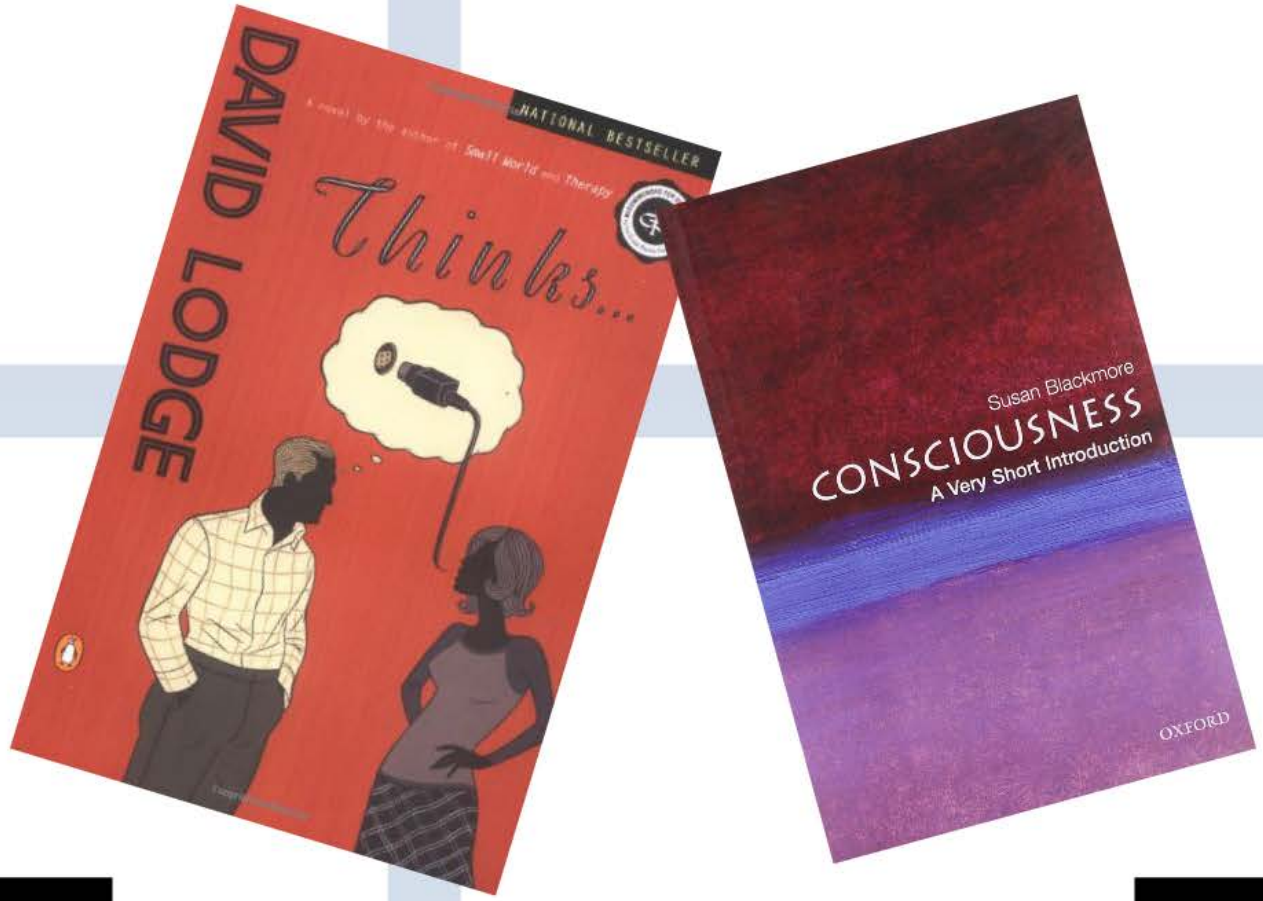
Conferences

No specific conference on Machine Consciousness!

- AGI: Artificial General Intelligence
- http://agi-society.org
- BICA: Biologically Inspired Cognitive Architectures
- http://bicasociety.org
- ASSC Conference:
- http://www.theassc.org
- Towards a Science of Consciousness:
- http://www.consciousness.arizona.edu

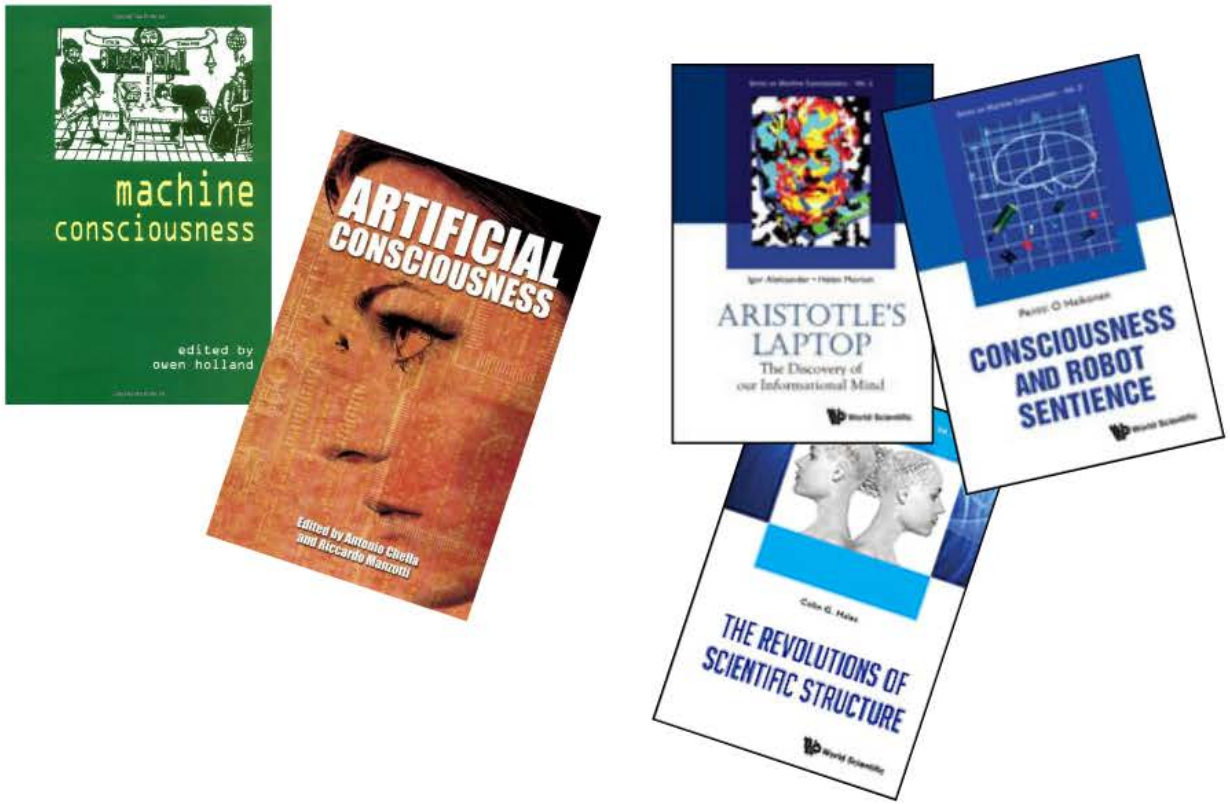
Intro

- D. Lodge: Thinks..., Penguin, 2002
- S. Blackmore: Consciousness, A Very Short Introduction. Oxford University Press, 2005



Machine Consciousness

- O. Holland (ed.): Machine Consciousness, Imprint Academic, 2003
- A. Chella, R. Manzotti (eds.): Artificial Consciousness, Imprint Academic, 2007
- Book Series on Machine Consciousness:
- http://www.worldscientific.com/series/smc



Websites

- http://www.conscious-robots.com/ (thanks to Raul Arrabales)
- Association for the Scientific Study of Consciousness:
- http://www.theassc.org/
- Center for Consciousness Studies -
- University of Arizona:
- http://www.consciousness.arizona.edu/

Journals

- International Journal of Machine Consciousness (currently discontinued)
- http://www.worldscinet.com/ijmc/
- Journal of Consciousness Studies
- http://www.imprint.co.uk/jcs.html
- Consciousness and Cognition
- www.elsevier.com/locate/concog

Conferences

No specific conference on Machine Consciousness!

- **AGI: Artificial General Intelligence**
- **http://agi-society.org**
- **BICA: Biologically Inspired Cognitive Architectures**
- **http://bicasociety.org**
- **ASSC Conference:**
- **http://www.theassc.org**
- **Towards a Science of Consciousness:**
- **http://www.consciousness.arizona.edu**

Thank you for your "conscious" attention!

Thanks to Jorge Cham!

http://roboticslab.dicgim.unipa.it