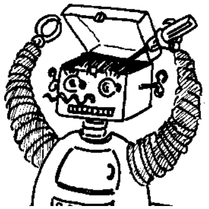
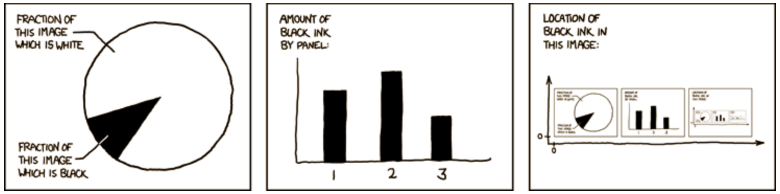


http://xkcd.com/


FRACTION OF THIS IMAGE WHICH IS WHITE
FRACTION OF THIS IMAGE WHICH IS BLACK
AMOUNT OF BLACK INK BY PANEL:
1
2
3
LOCATION OF BLACK INK IN THIS IMAGE:
0

Maryland Metacognition Seminar

TOWARD ROBOT CONSCIOUSNESS

ANTONIO CHELLA

University of Palermo

Motivations for MC

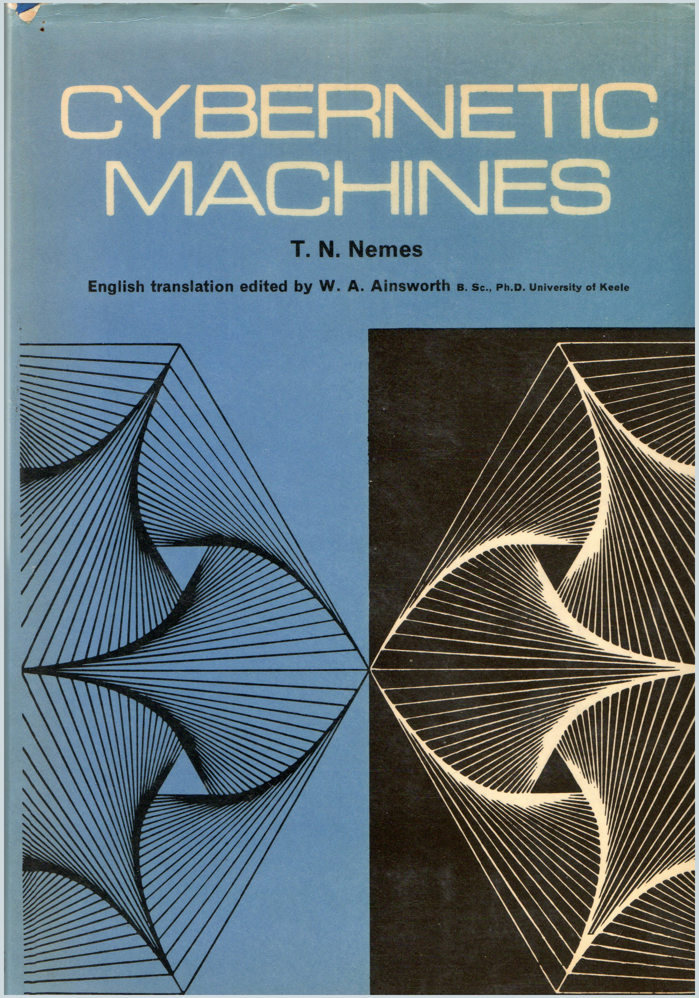
- Artefacts like us: consciousness, emotion and affect, experience, imagination, creativity (Robotics)
- Studying natural systems with computer laboratory models (Cognitive Science)
- Proficient machines (Intelligent Control)

When machines will be conscious?

- A conscious machine could require the same complexity of the human brain
- We could be able to build such a machine in 2029
- Human Brain Project (EU Flagship Project)
- Randal Koene *Carboncopies*
- ….

A Brief History

- Nemes: Cybernetic Machines
- Published in Budapest in 1962
- Translated and published in English in 1970



CYBERNETIC MACHINES

T. N. Nemes

English translation edited by W. A. Ainsworth B. Sc., Ph.D. University of Keele

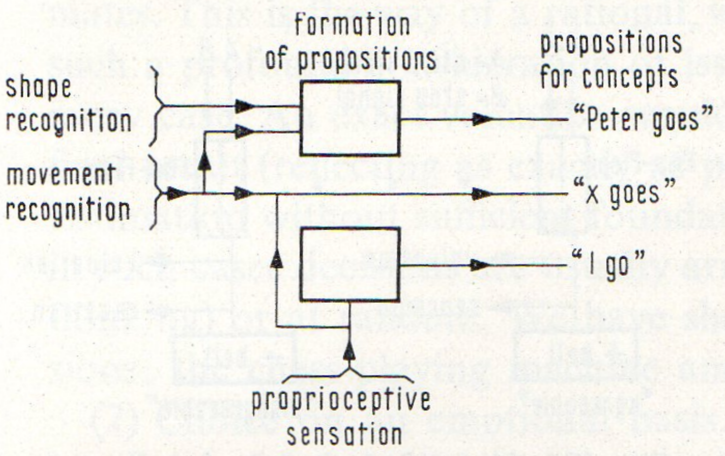
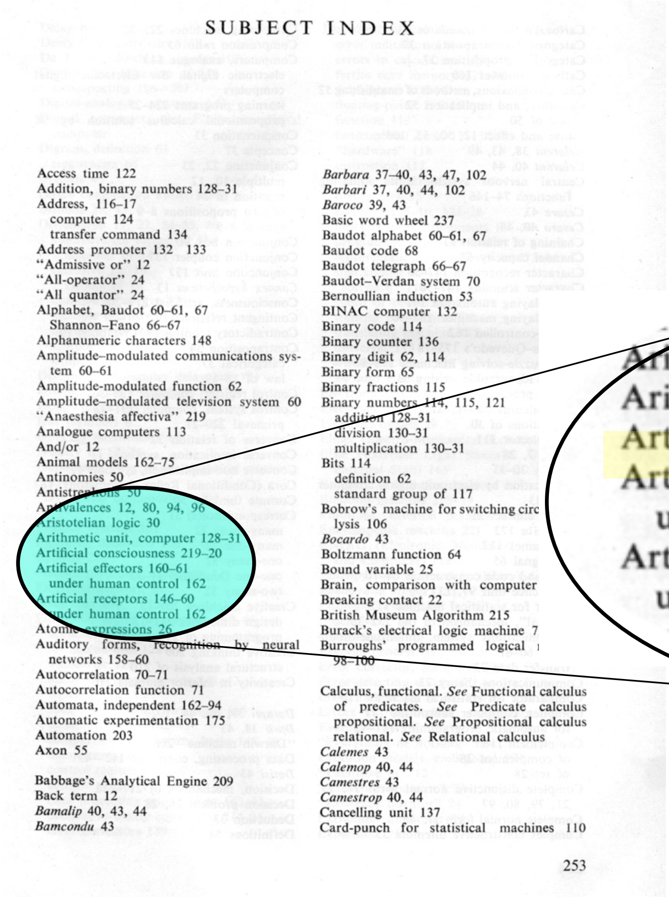
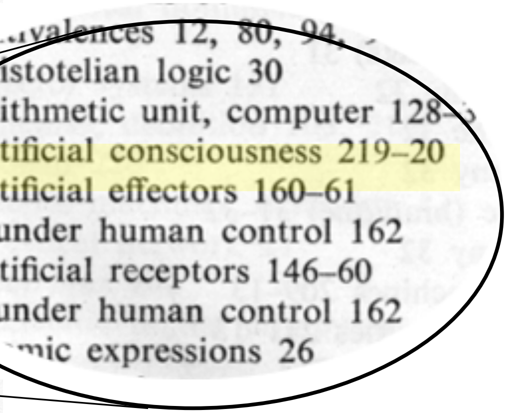


Fig. 137. A preliminary conceptual sketch of the "first person"



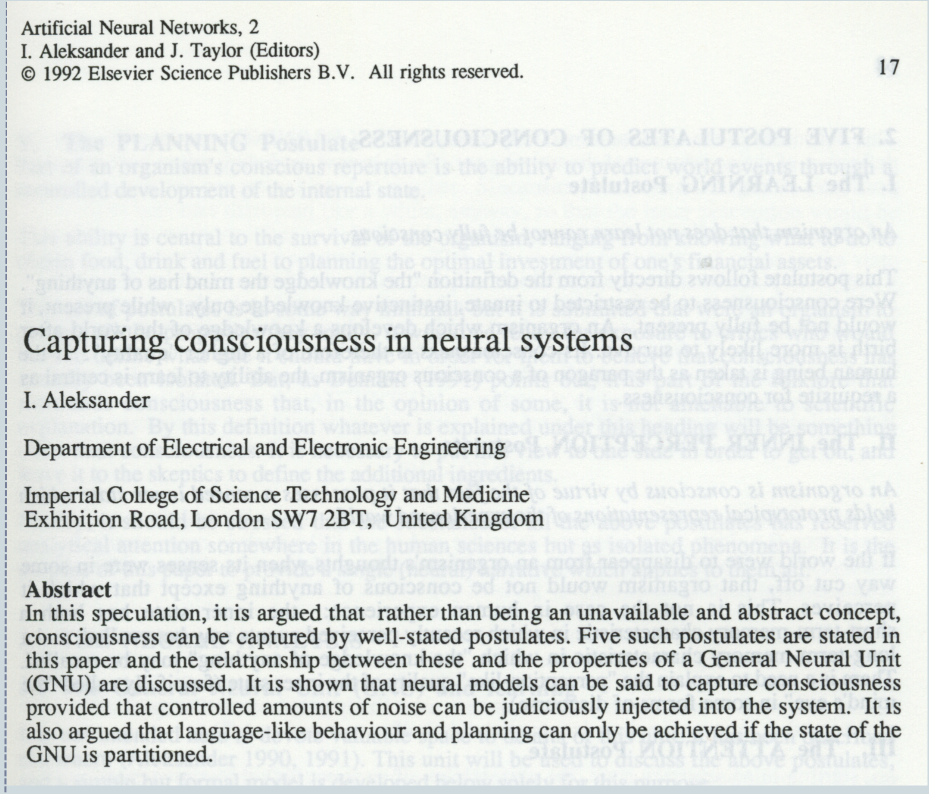
SUBJECT INDEX



ivalences 12, 80, 94,
Aristotelian logic 30
Arithmetic unit, computer 128–
Artificial consciousness 219–20
Artificial effectors 160–61
  under human control 162
Artificial receptors 146–60
  under human control 162
mic expressions 26

1992: Igor Aleksander ICANN Brighton

- Modern view of machine consciousness
- "The hunting season of machine consciousness is open"



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.

2001: Cold Spring Harbour Meeting
Can a machine be conscious

“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 in final report)

ieee spectrum

THE MAGAZINE OF TECHNOLOGY INSIDERS

06.08

WILL SUPERSM
MACHINES LET
US LIVE FOREV
OR RENDER
US OBSOLETE?

THE RA

RTIFICIAL
NSCIOUSNESS

Antoni
cardo M

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

WEEKLY April 3 - 9, 2010

.95 CAN$5.95 No2754

International Journal of
MACHINE
CONSCIOUSNESS

Volume 1 · Number 1 · June 2009

Editor-in-Chief

Antonio Chella
University of Palermo, Italy

World Scientific

ELSEVIER

ARTIFICIAL INTELLI
MEDICINE

Volume 44 No 2 October 2008 ISSN 0933-3657

Special Issue:
Artificial Consciousness

Guest Editors:
Giorgio Buttazzo and Riccardo Manzotti

machine
consciousness

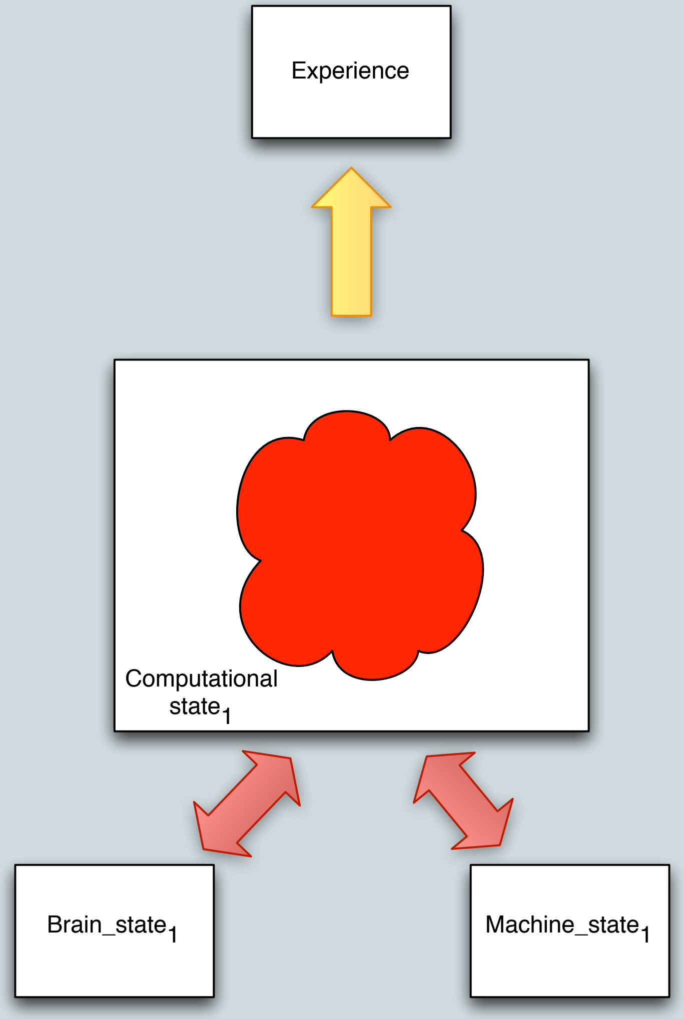
edited by
owen holland

EXPERIENCE OF REDNESS

S

Consciousness and Computational Mind

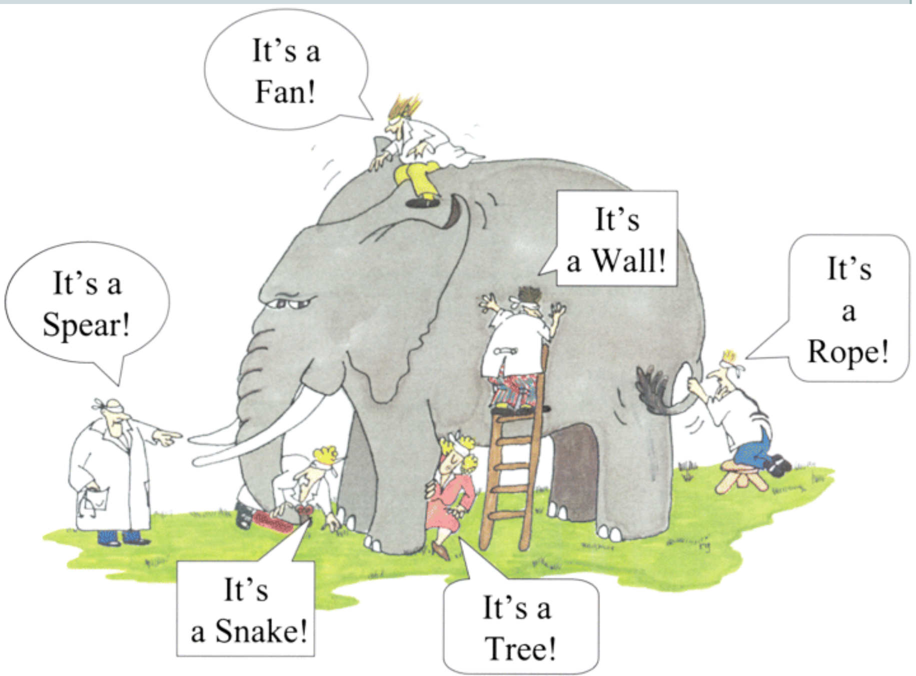
- The elements of the conscious awareness are caused by information and processes of the computational mind that:
  - are active
  - have other privileged properties



Jackendoff 1987

Computational models for MC

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



http://www.nature.com/ki/journal/v62/n5/fig_tab/4493262f1.html

Information Integration Theory (Tononi)

- Conscious experience is differentiated
  - the potential repertoire of different conscious states is huge
- Conscious experience is integrated
  - every conscious state is experienced as a single entity
- The substrate of conscious experience must be an integrated entity able to differentate among an enormously big repertoire of different states

Information

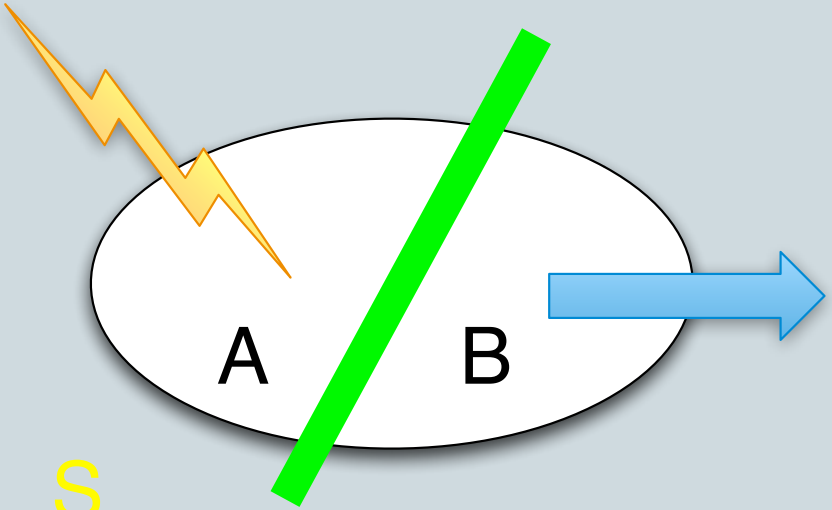
- Galileo and a photodiode in front of a flashing screen
- The same answers!
- But... Galileo is able to discriminate among a **huge** number of states
- How much information is generated:
- Entropy $H = -\sum p_i \log_2 p_i$

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!

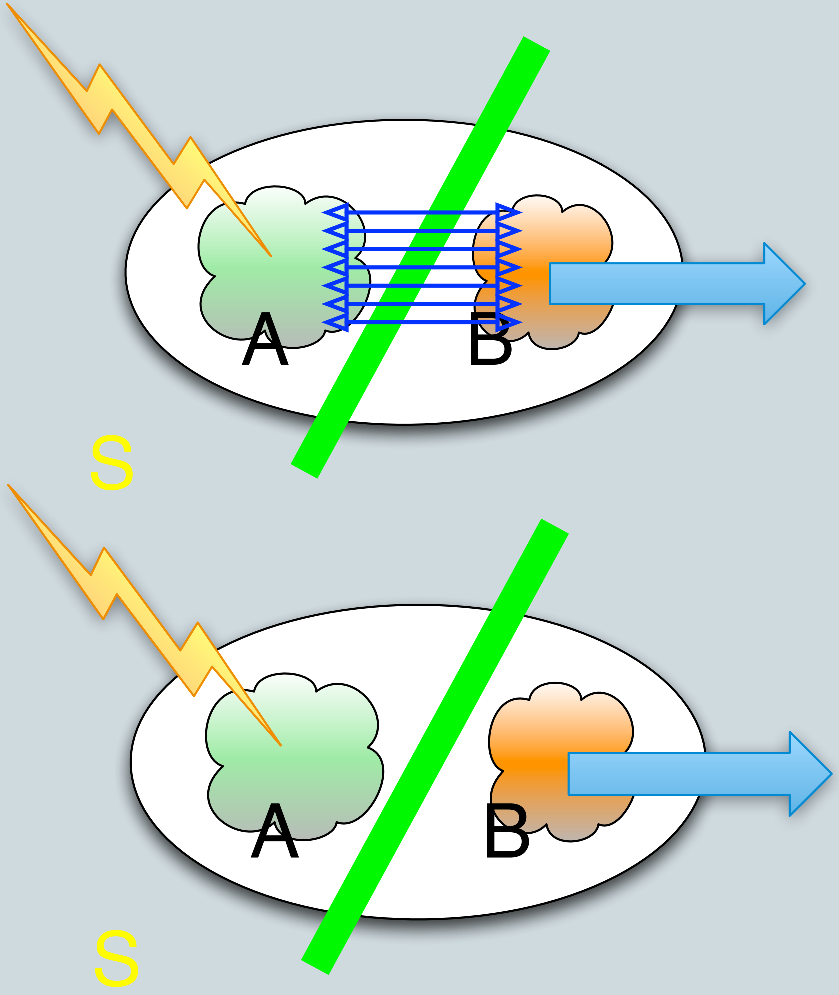
Effective Information EI

- 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)



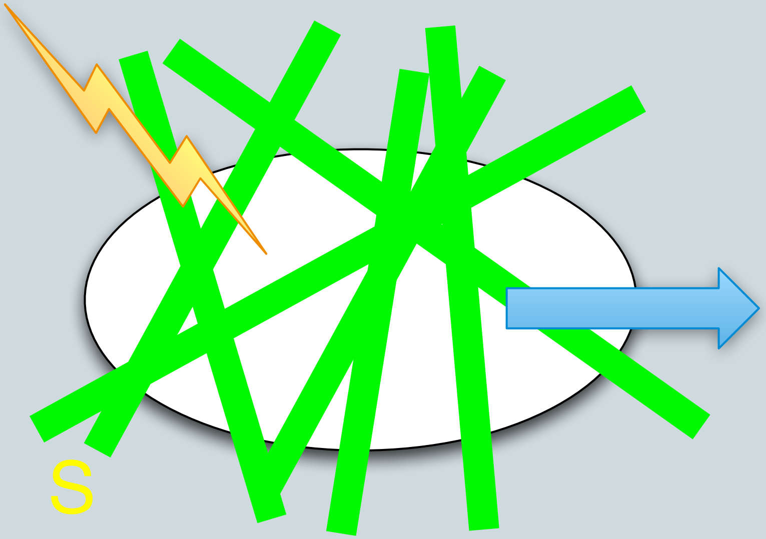
Information Integration

- The system S can integrate information only if A and B are higly 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 for 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
- $\Phi(S) = EI\left({}^{MIB}A \rightleftharpoons B\right)$
- Complex: a subset of S with Φ>0 not included in larger subsets with higher Φ



Φ and Consciousness

- A conscious complex is a complex with high Φ(S)
- “Complexes are the subjects of experience, being the locus where information can be integrated”
- Consciousness is not an all-or-none but graded by Φ(S)
- Complex contributes to conscious experience, the other parts of the systems do not, even if they are connected to it
- Experience, e.g., information integration, is a fundamental quantity as mass, charge, energy

Φ and Machine Consciousness

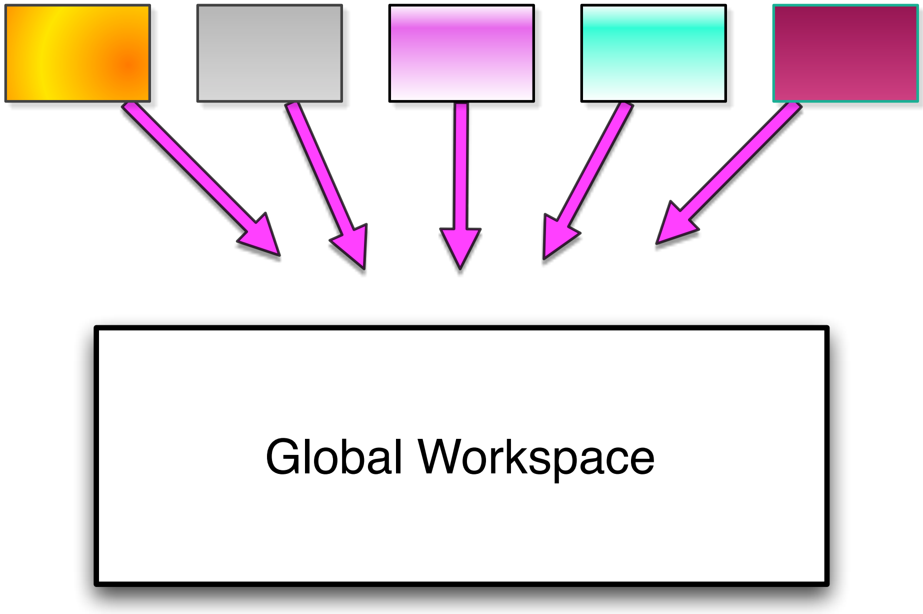
- Any physical system have subjective experience to the extent that it is capable to integrate information
- It could be possible in principle to build conscious artifact by endowing them with a complex of high Φ(S)
- A conscious vision machine should be able to differentiate the key features of a scene from the immense range of possible scenes and to integrate them in a detailed description of the scene itself

Global Workspace Theory (Baars)

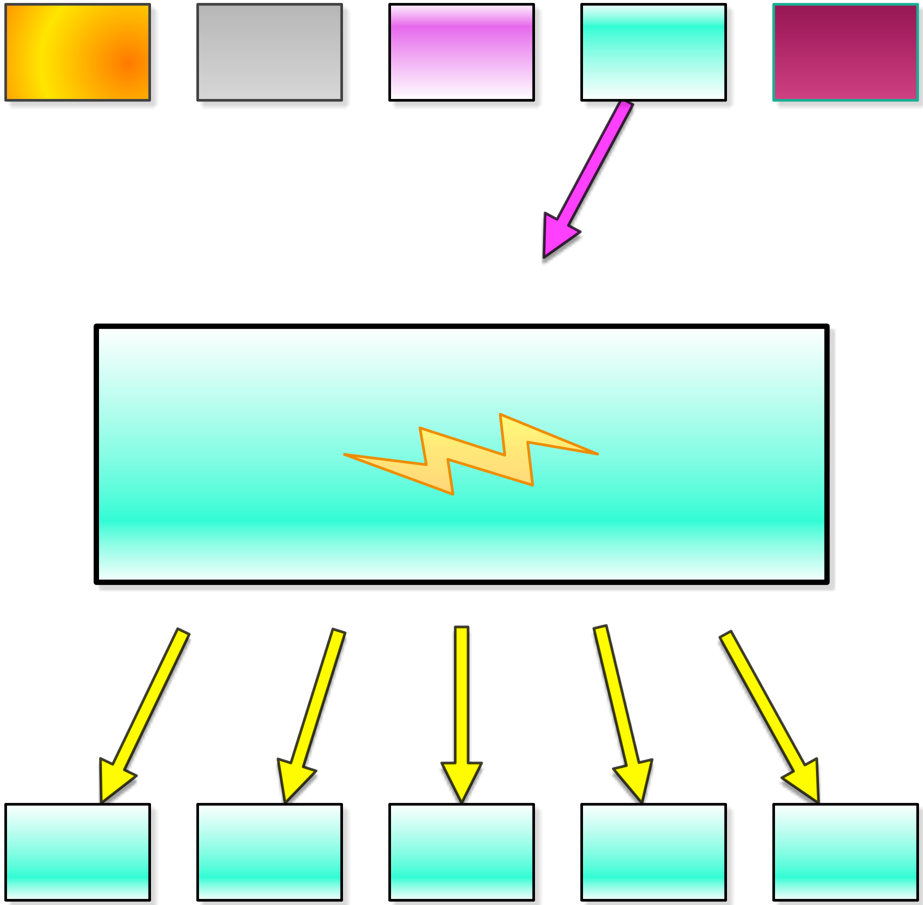
- The brain is a collection of unconscious specialized processors
- Consciousness is serial with limited capacity
- Consciousness is associated with a *global workspace* whose contents "broadcast" to many processors

- Contexts shape conscious contents
- Contexts may work together to constraint conscious events
- Motives and emotions are parts of goal contexts

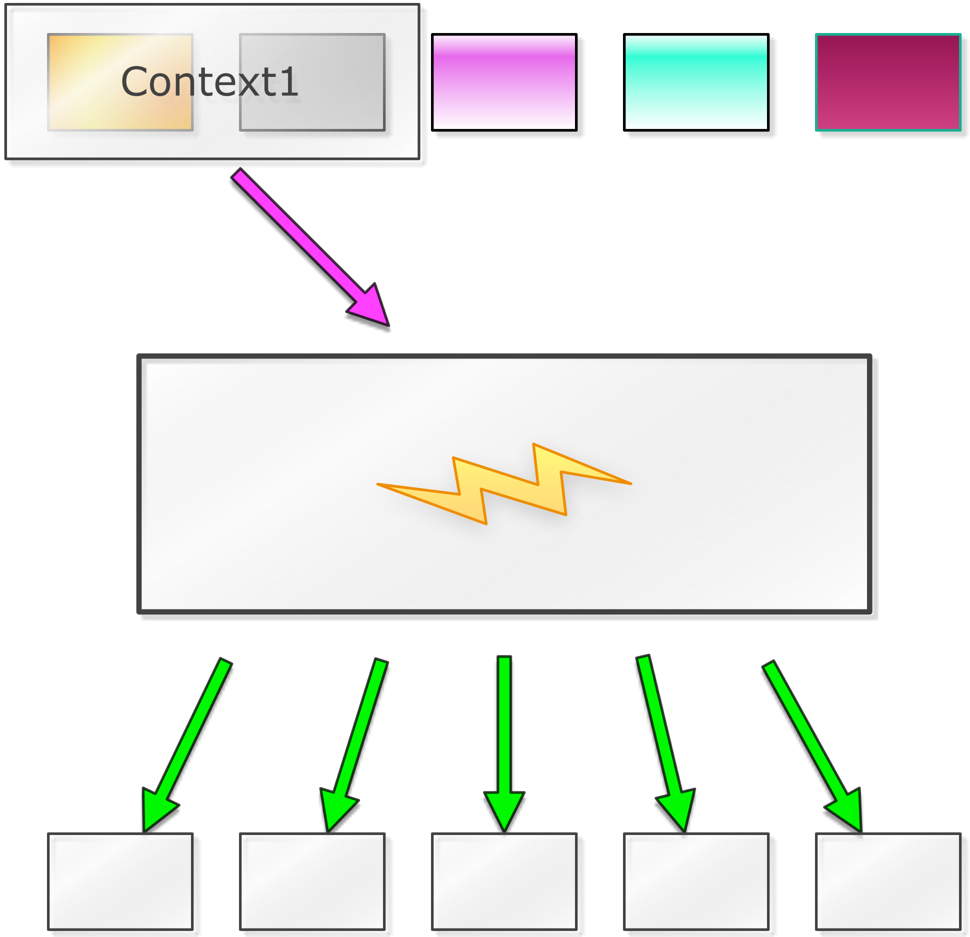
- *Self* is the most general unifying context



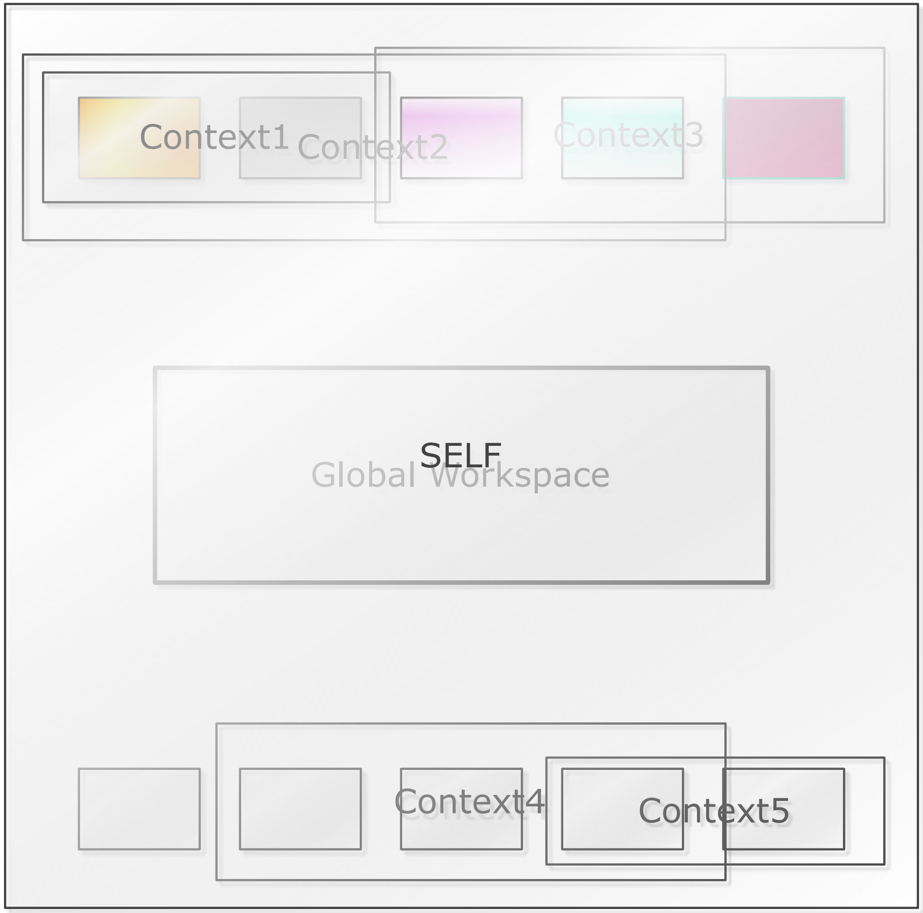
Several unconscious processors compete for access to GW in order to recruit more processors



The winner processor gain GW, i.e., consciousness and broadcast to the other processors



Context may allow for a coalition of processors in order to shape the content of consciousness



Self is the deepest level of context: the basic intentions and expectatons we have towards the world, ourself and each others (Baars)

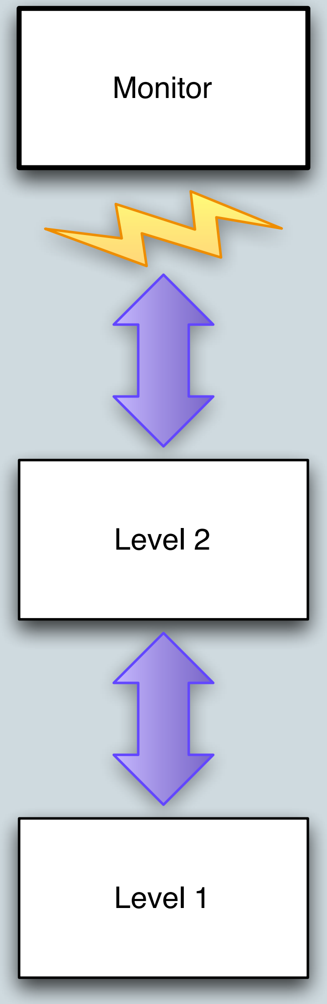
GWT Implementations

- LIDA (Franklin et al.)
- Shanahan’s Cognitive Architecture
- Dehaene’s Neuronal Workspace Model
- CERA – CRANIUM (Arrabales et al.)

0:00 / 1:52

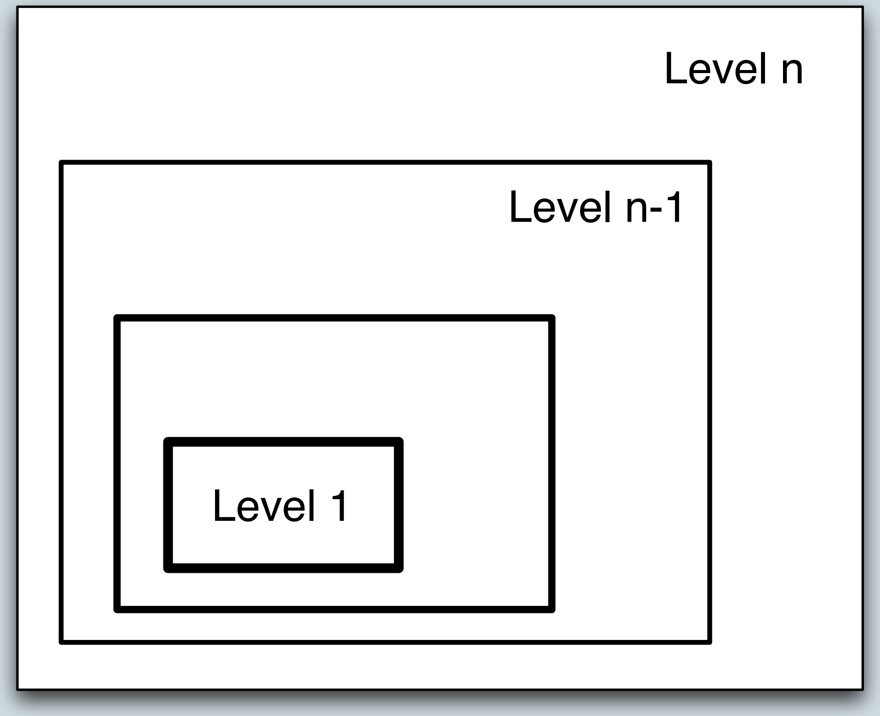
Introspection/monitoring models

- 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



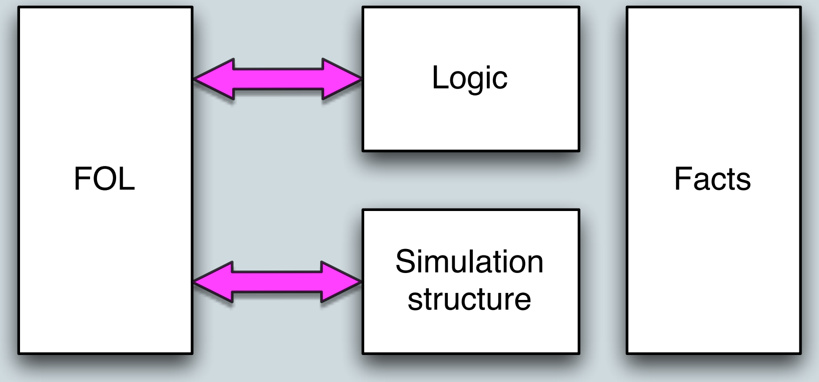
Recursion models

- Recursion of modules in the computational mind
- Level n comprises level (n-1)
- Introspection, self reflection modules



FOL (Weyhrauch)

- “Mechanized” formal reasoning
- Simulation structure: the interpretation model
- Association of analogue representation to the symbolic formalism
- Exploiting meta-level representations
- Reflection about the system itself and its own capabilities

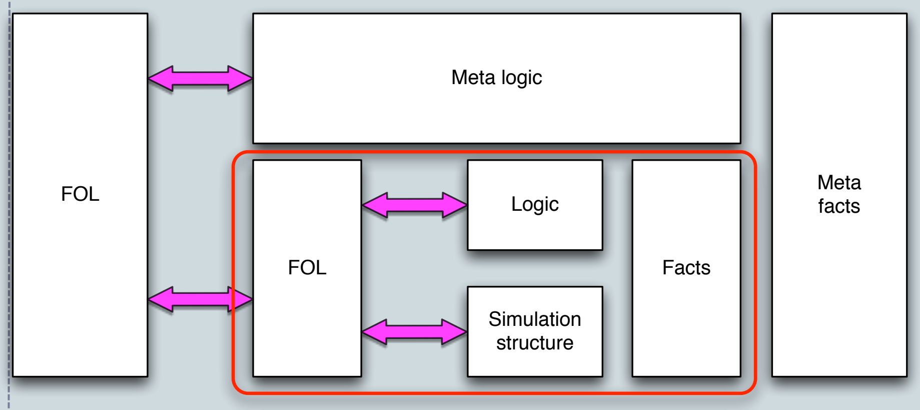


FOL

Logic

Simulation structure

Facts



FOL

Meta logic

FOL

Logic

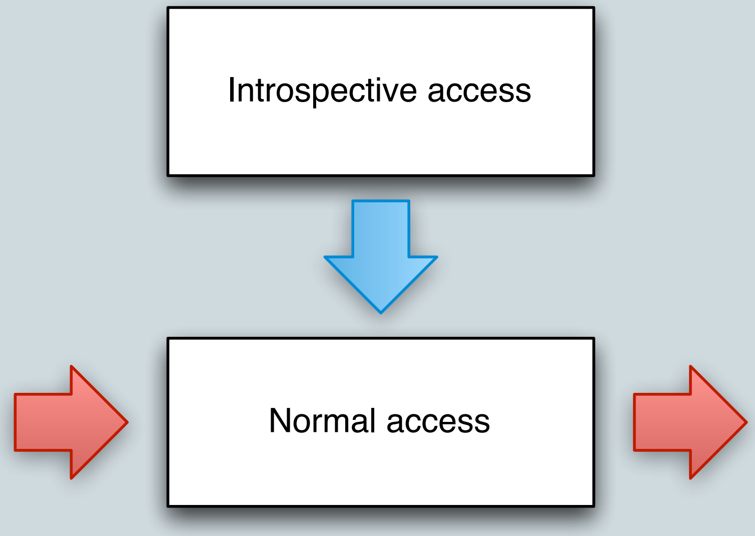
Simulation structure

Facts

Meta facts

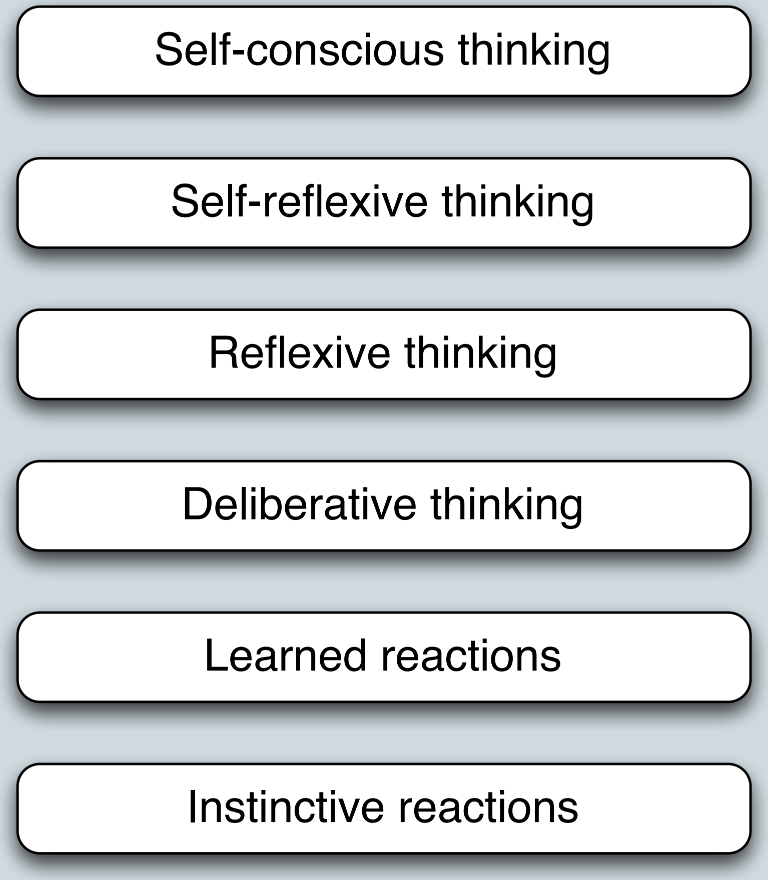
Self-model (McDermott)

- Normal access: information about the world
- Introspective access: information about the robot itself
- Self-model: S=M
- Robot with a model of itself



Layers of reflection (Minsky)

- Multiple layers of critics
- Each layer reflects and critics upon the layers beneath
- Capabilities, limitations and improvements
- EM-ONE (Singh)



Self-conscious thinking

Self-reflexive thinking

Reflexive thinking

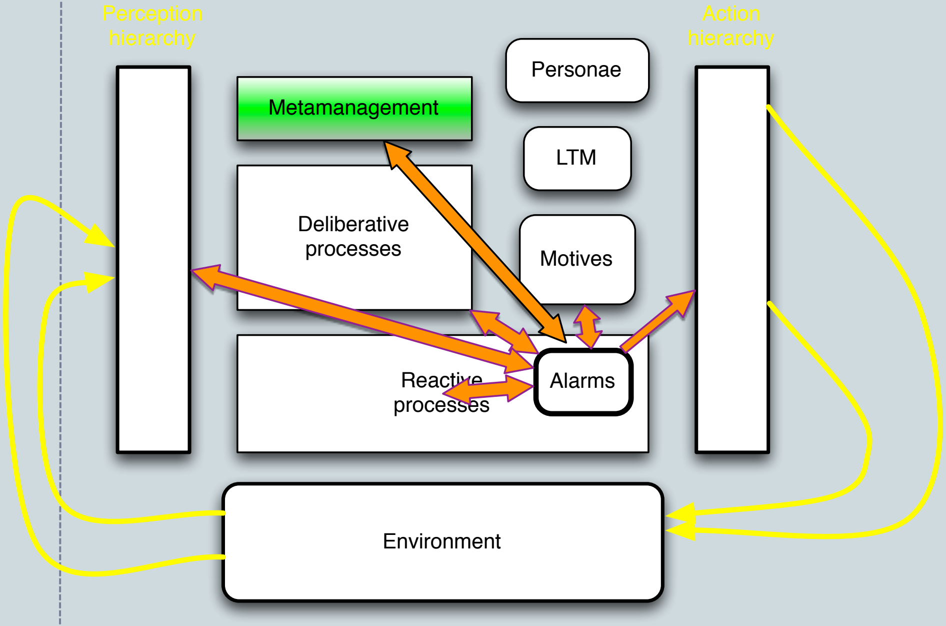
Deliberative thinking

Learned reactions

Instinctive reactions

Cog-Aff (Sloman)

- Three main levels
- A framework architecture
- Reactive mechanisms
- Deliberative reasoning
- Meta management (reflective processes)



Mental Situation Calculus (McCarthy)

- Introspective reasoning
- Propositions are of mental nature
- A robot may reason about its own mental states
- Situation calculus describes the evolution of robot mental states: knowledge, abilities, intentions, past history, …
- Introspection as problem solving by considering evolutions of mental states and not just evolution of the external word

Introspective knowledge

- Holds(Know(p), S_i)
- Holds(Know(Not(Know(p))), S_i)
- Holds(Know(Not(Know(Telephone(Mike)))), S_i)
- The robot may search for Mike’s telephone number in the phone book

Examples of mental actions

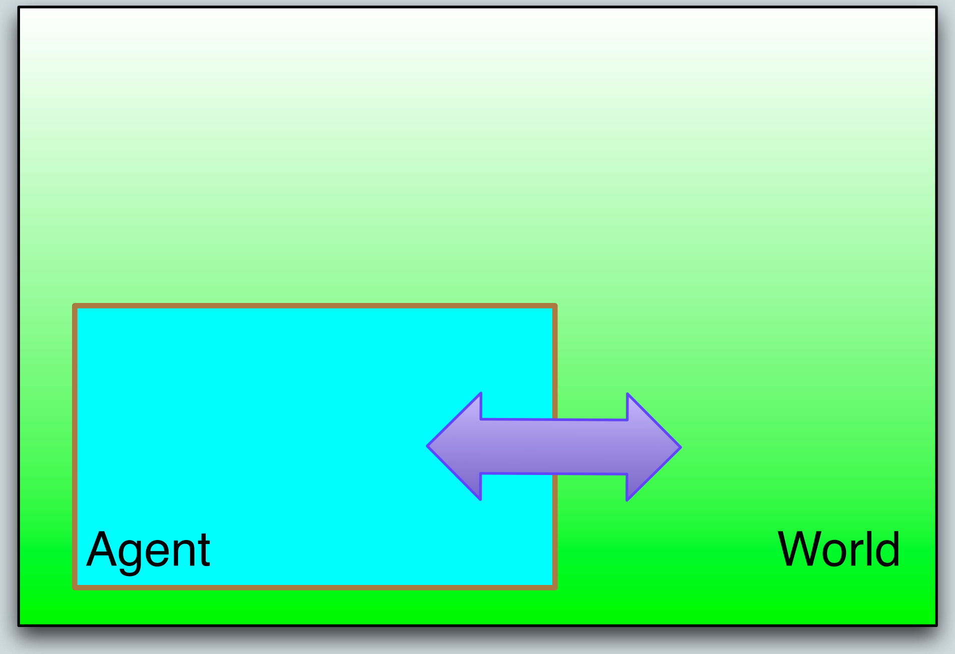
- Holds(Knows(p), Result(Learn(p), S_i))
- Occurs(Learn(p), S_i) → Holds(F(Know(p), S_i)
- After the learning action occurs, the robot will know p in the future

- Occurs(Forget(p), S_i) → Holds(F(Not(Know(p))), S_i)
- Occurs(foo, S_i) → Occurs(Forget(p), S_i)
- Forget is a side effect of some event foo

Internal models

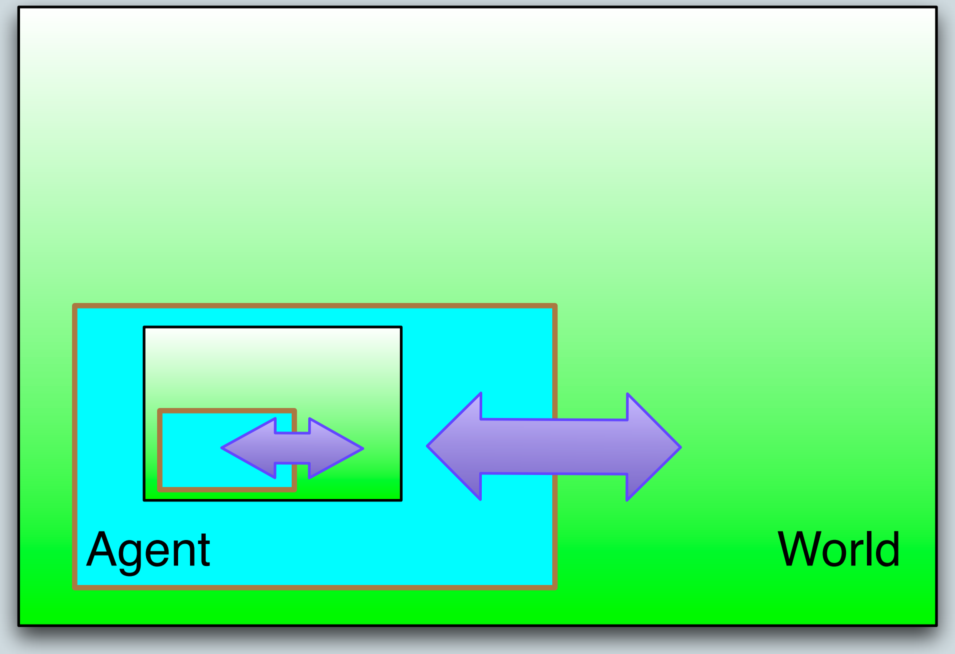
- An intelligent agent has an internal model of itself and of the external world
- Capability to simulate the external environment and the body actions
- Generation of expectations
- “Small scale model” of external reality (Craik)
- Popperian creatures (Dennett)

Agent and environment



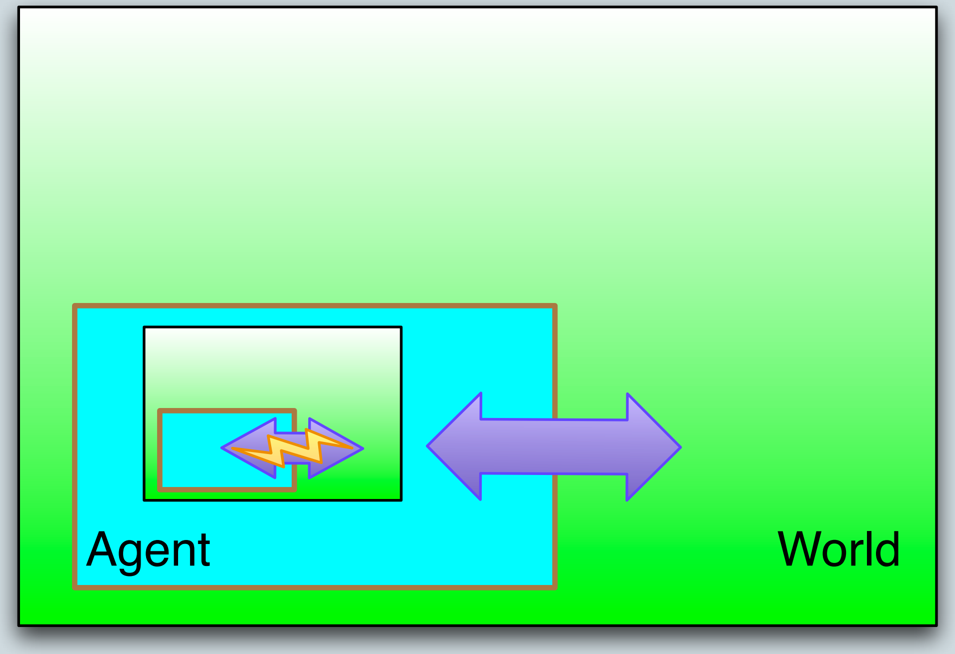
Holland 2007

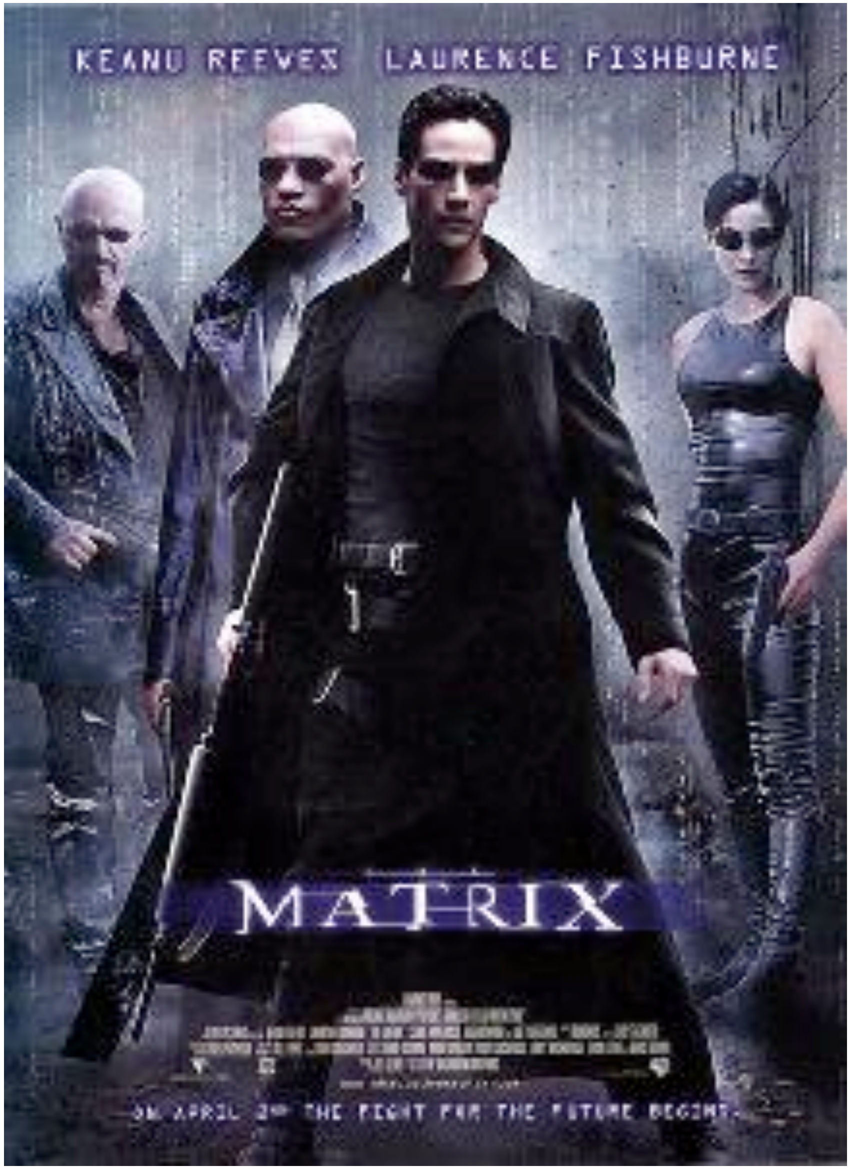
Agent with internal model of itself



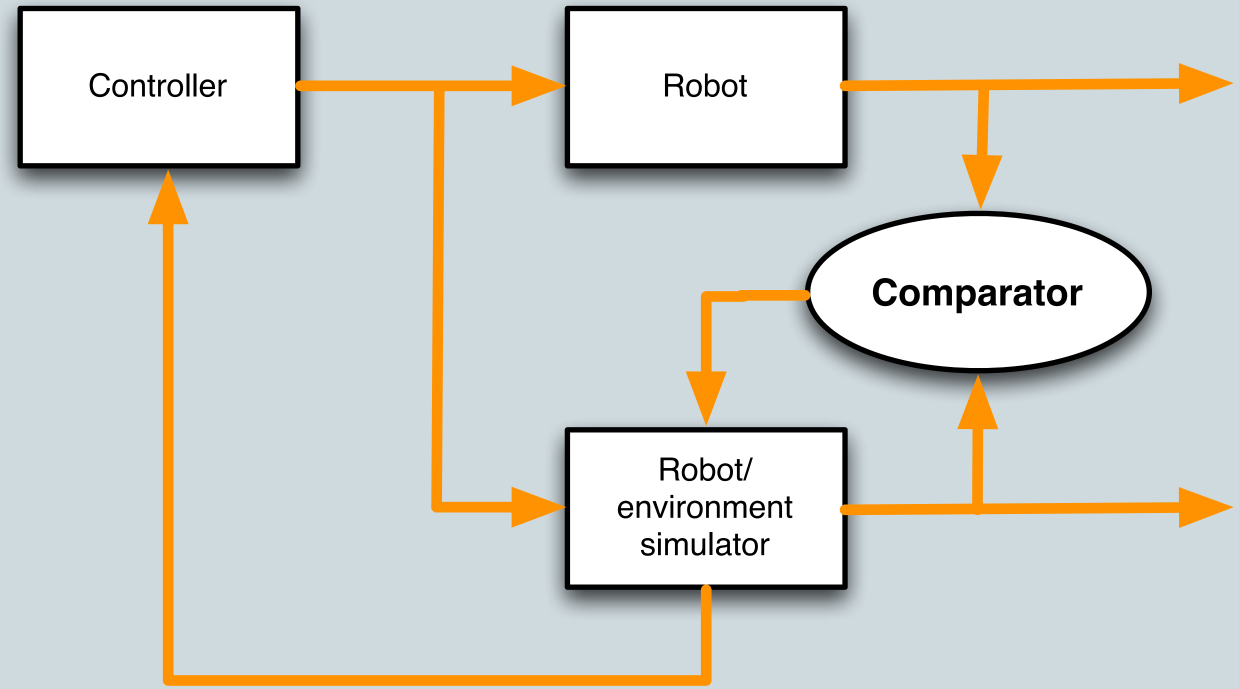
Internal model hypothesis

- Consciousness arises by the interaction between the internal model of the agent and the internal model of the environment



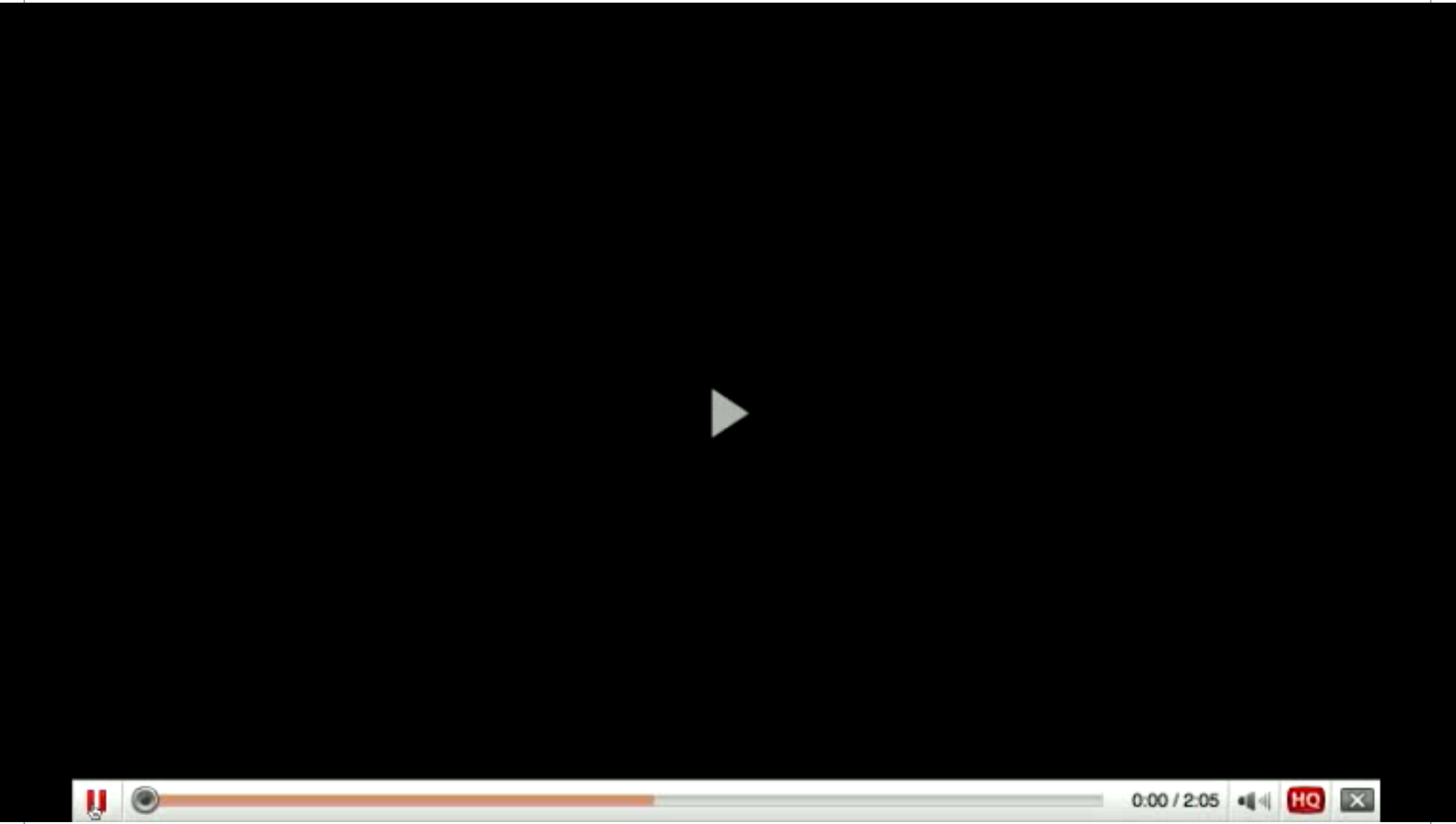

KEANU REEVES LAURENCE FISHBURNE
MATRIX
ON APRIL 2ND THE FIGHT FOR THE FUTURE BEGINS.

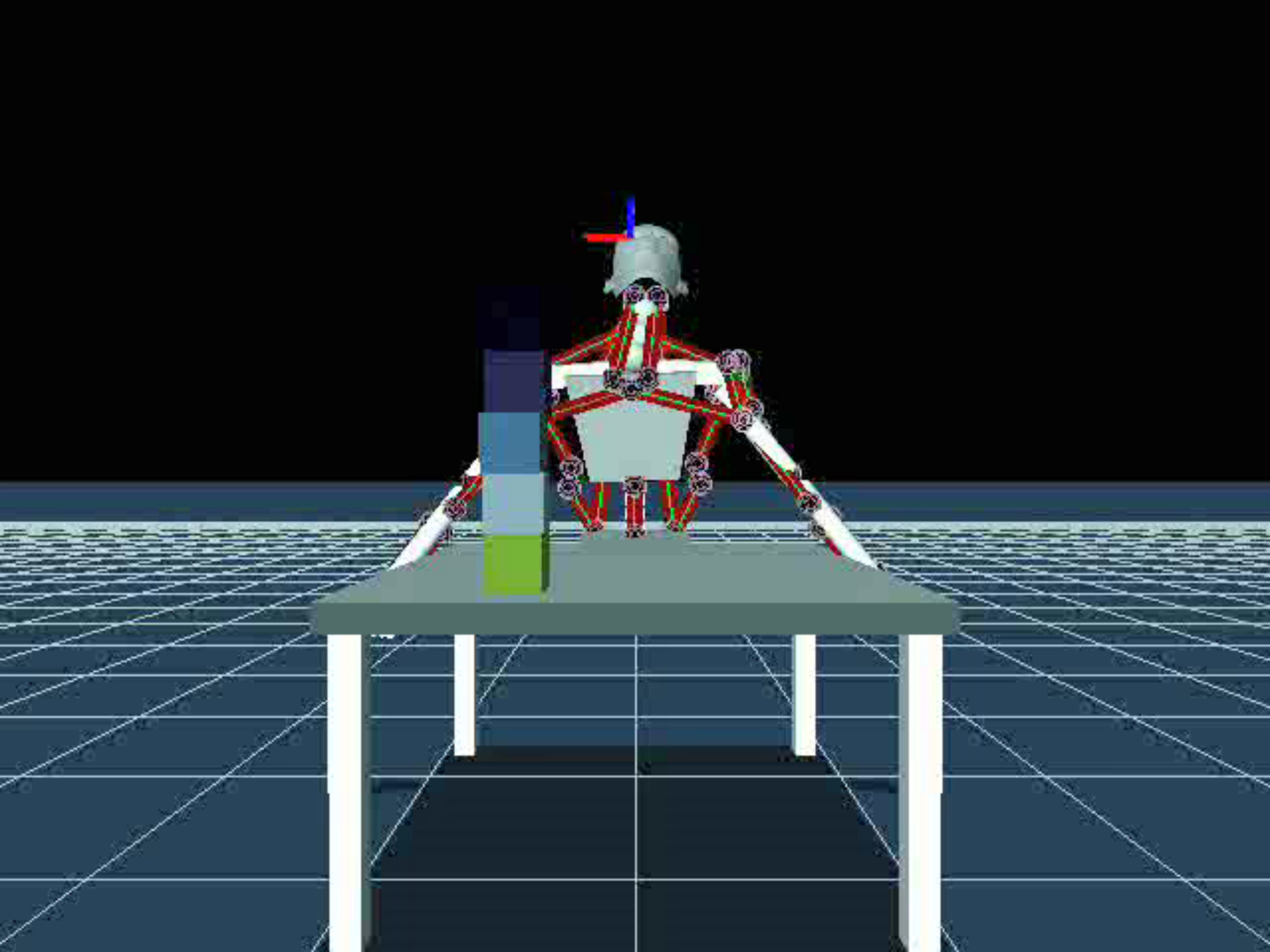
General Framework

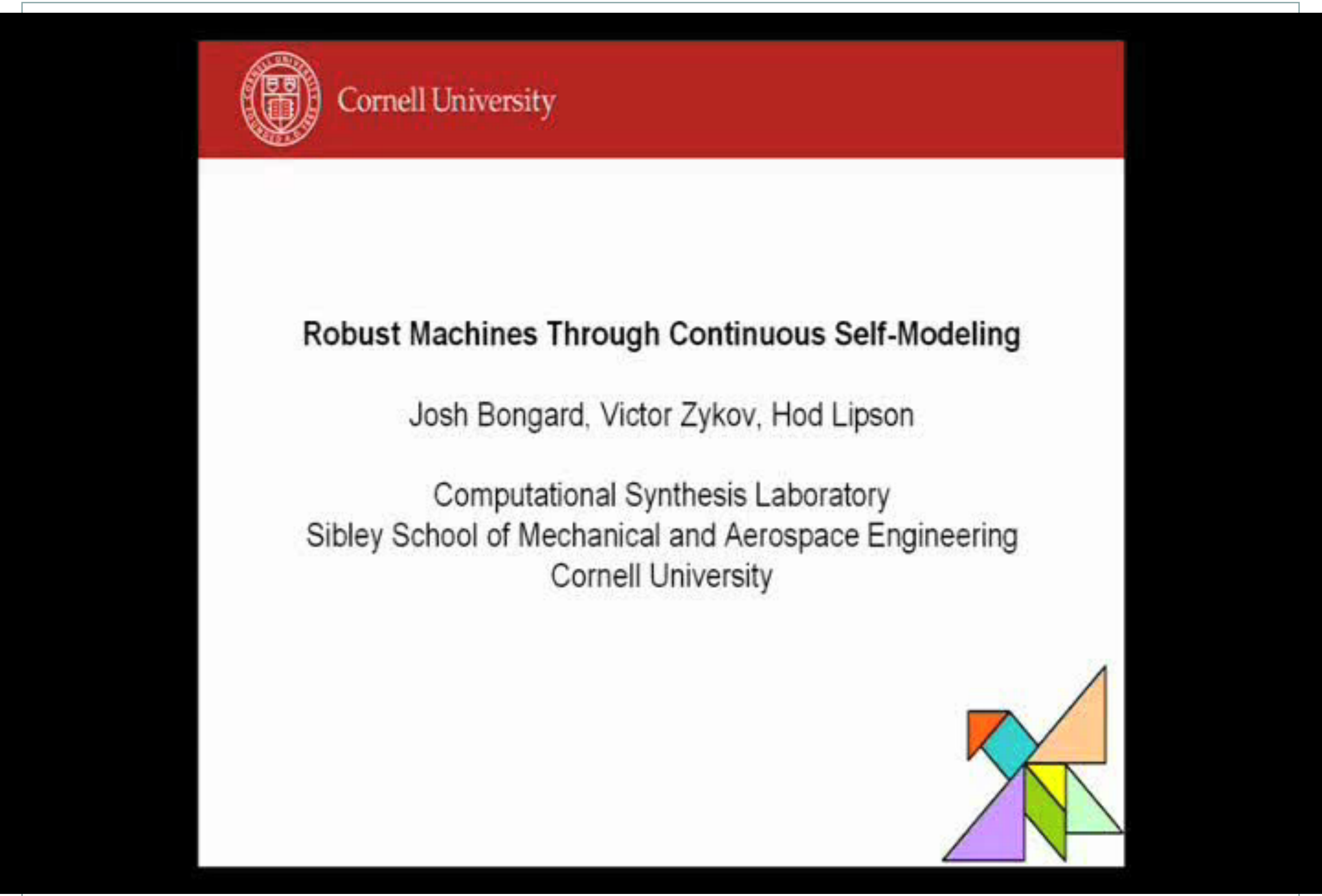


Implementations

- ECCEROBOT (Holland)
- Starfish robot (Bongard)
- Cicerobot (Chella)


0:00 / 2:05
HQ





Cornell University

Robust Machines Through Continuous Self-Modeling

Josh Bongard, Victor Zykov, Hod Lipson

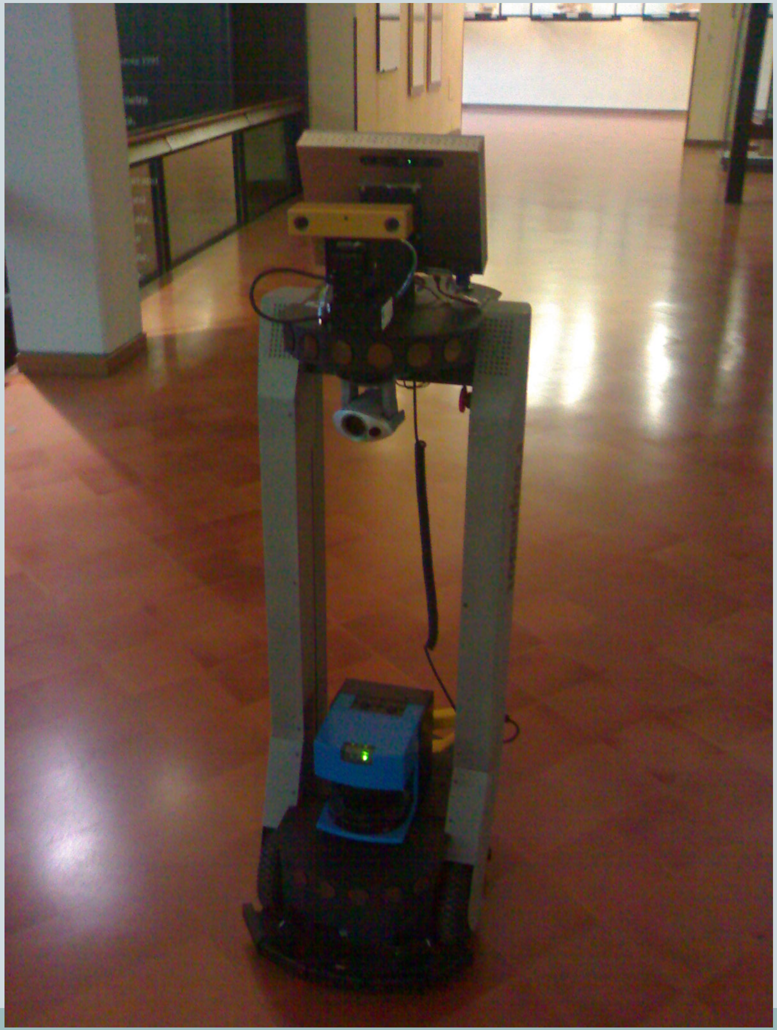
Computational Synthesis Laboratory
Sibley School of Mechanical and Aerospace Engineering
Cornell University

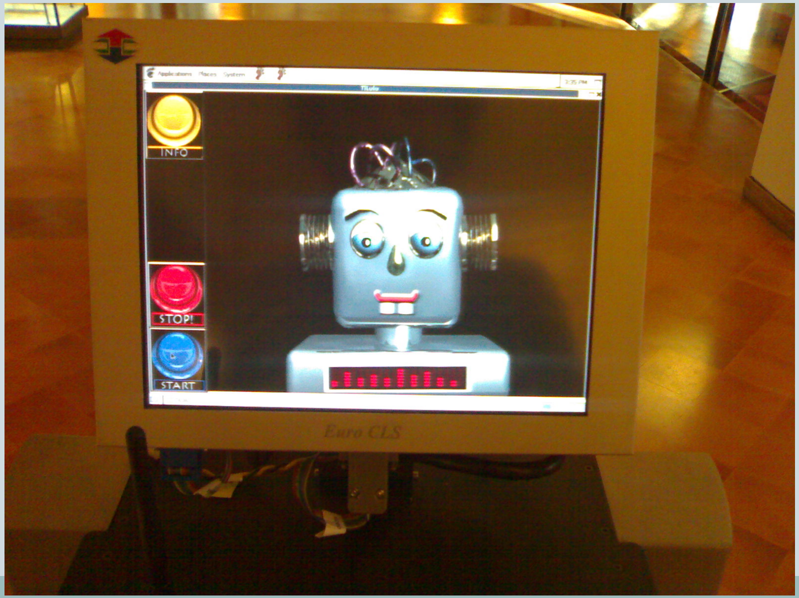
The Cicerobot Project

- Cicerobot is a museum guide robot tested in the Archeological Museum of Agrigento.
- Cicerobot allows default and interactive tours
- The user can introduce preferences to plan an ad hoc tour.
- Robot platforms equipped with a pan-tilt stereo head, laser range-finder and sonars.

Cicerobot

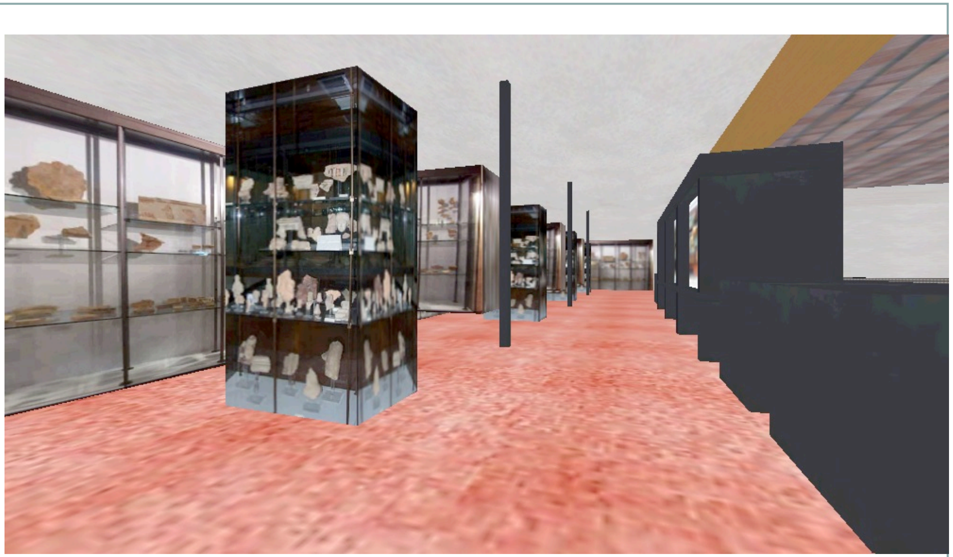
Cicerobot II



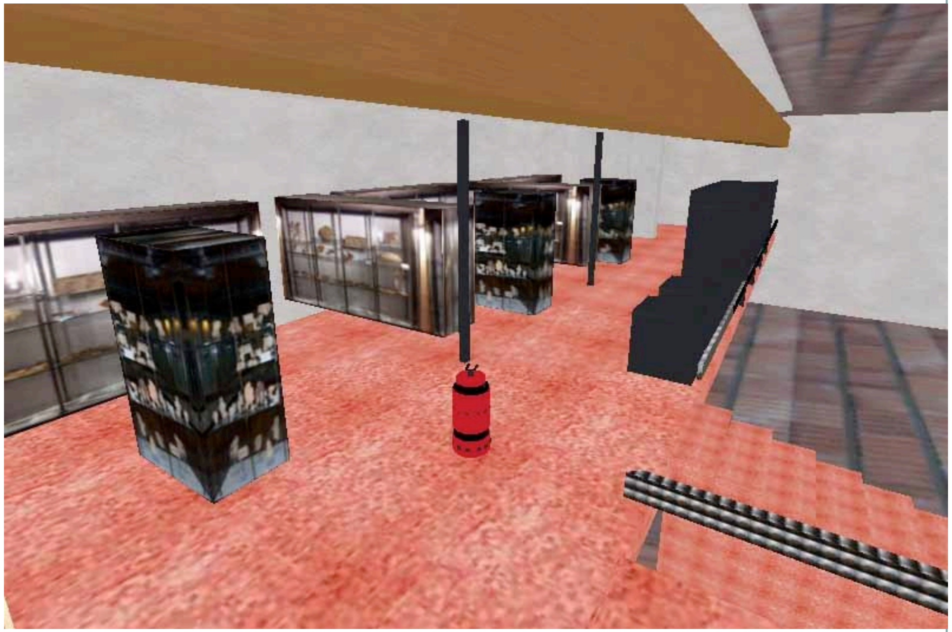


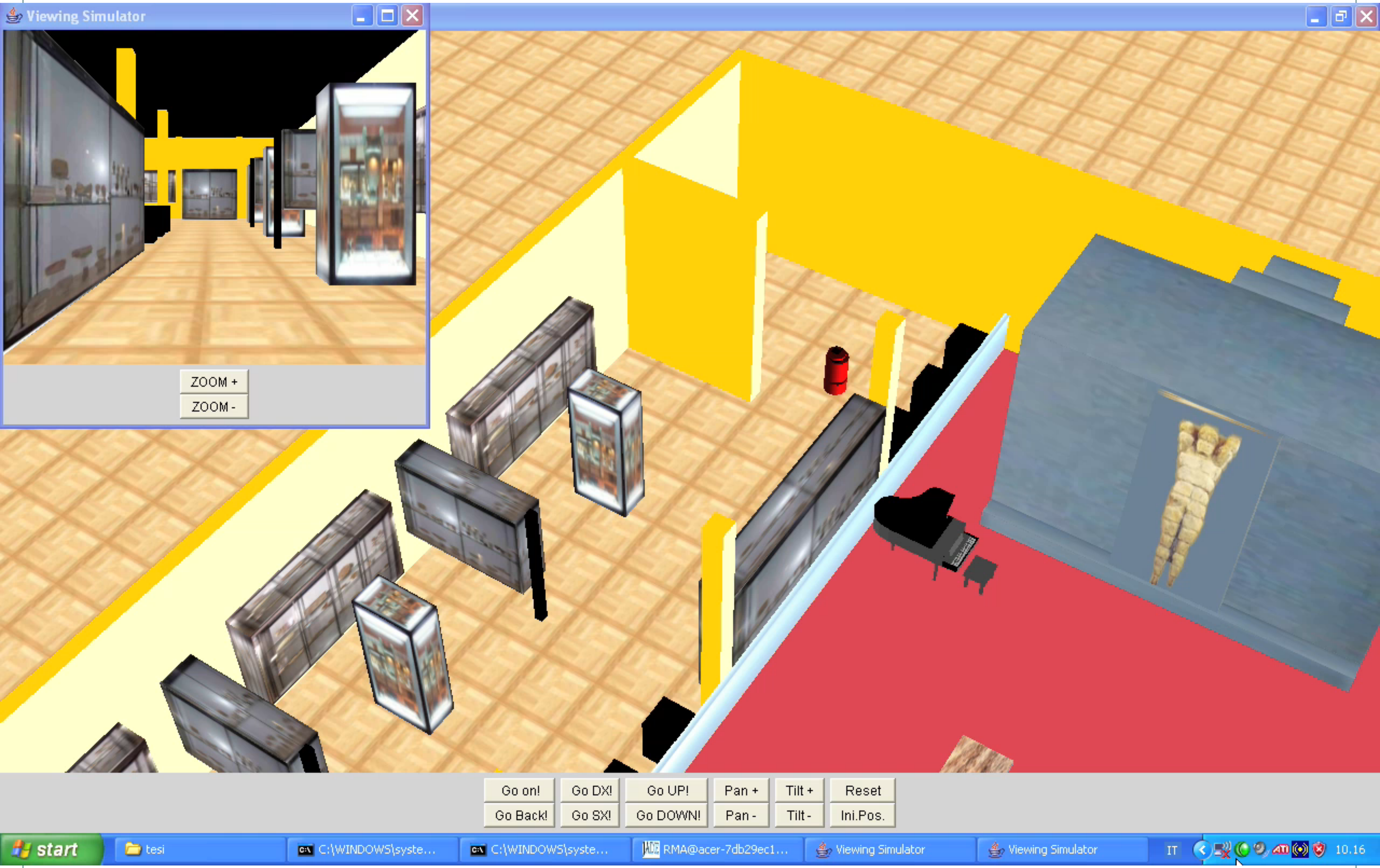
Robotanic

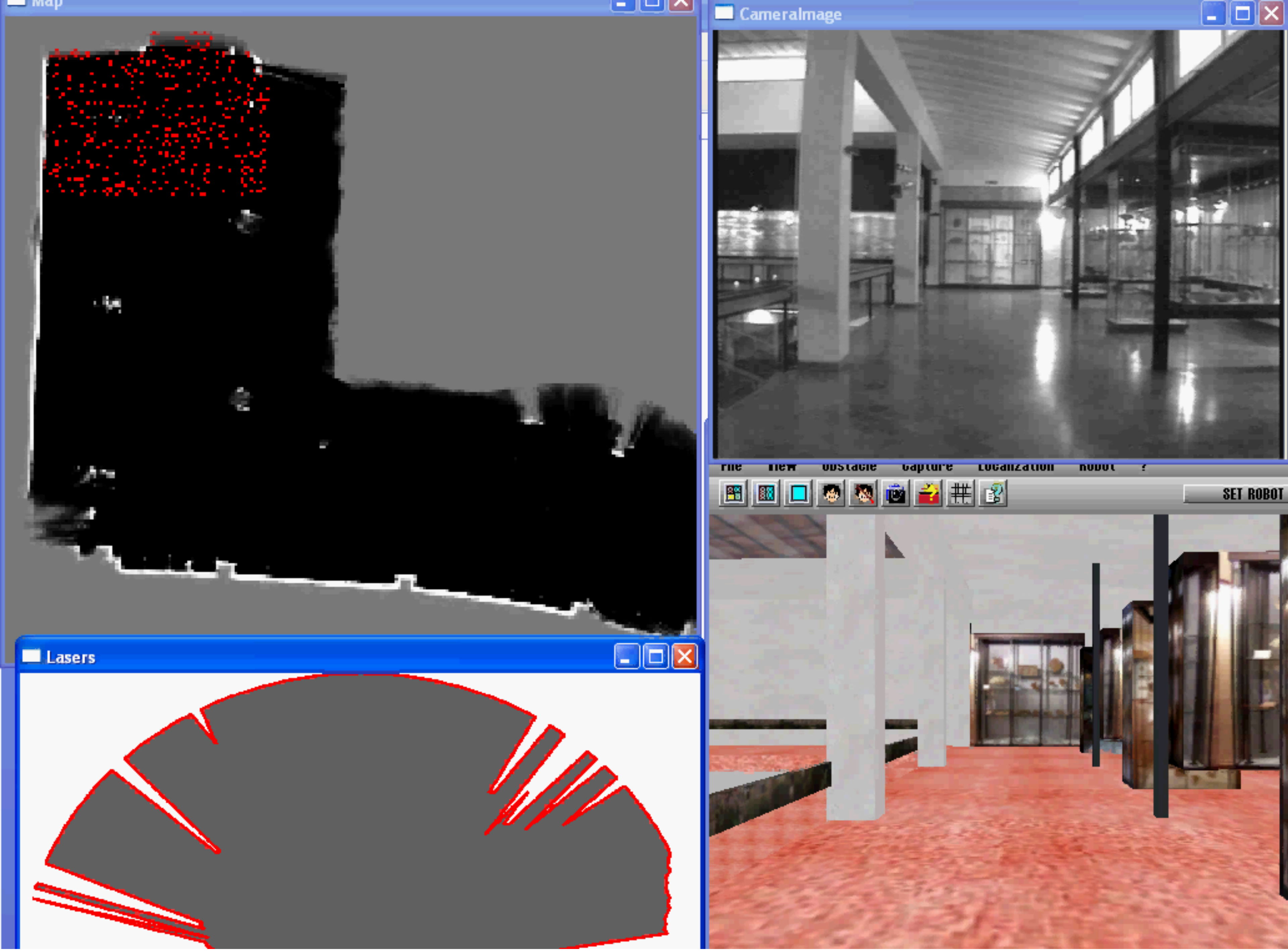
First order perceptions
(egocentric view)



Second order perceptions
(allocentric view)




Viewing Simulator
ZOOM +
ZOOM -
Go on!
Go Back!
Go DX!
Go SX!
Go UP!
Go DOWN!
Pan +
Pan -
Tilt +
Tilt -
Reset
Ini.Pos.
start
tesi
C:\WINDOWS\syste...
C:\WINDOWS\syste...
RMA@acer-7db29ec1...
Viewing Simulator
Viewing Simulator
IT
10.16

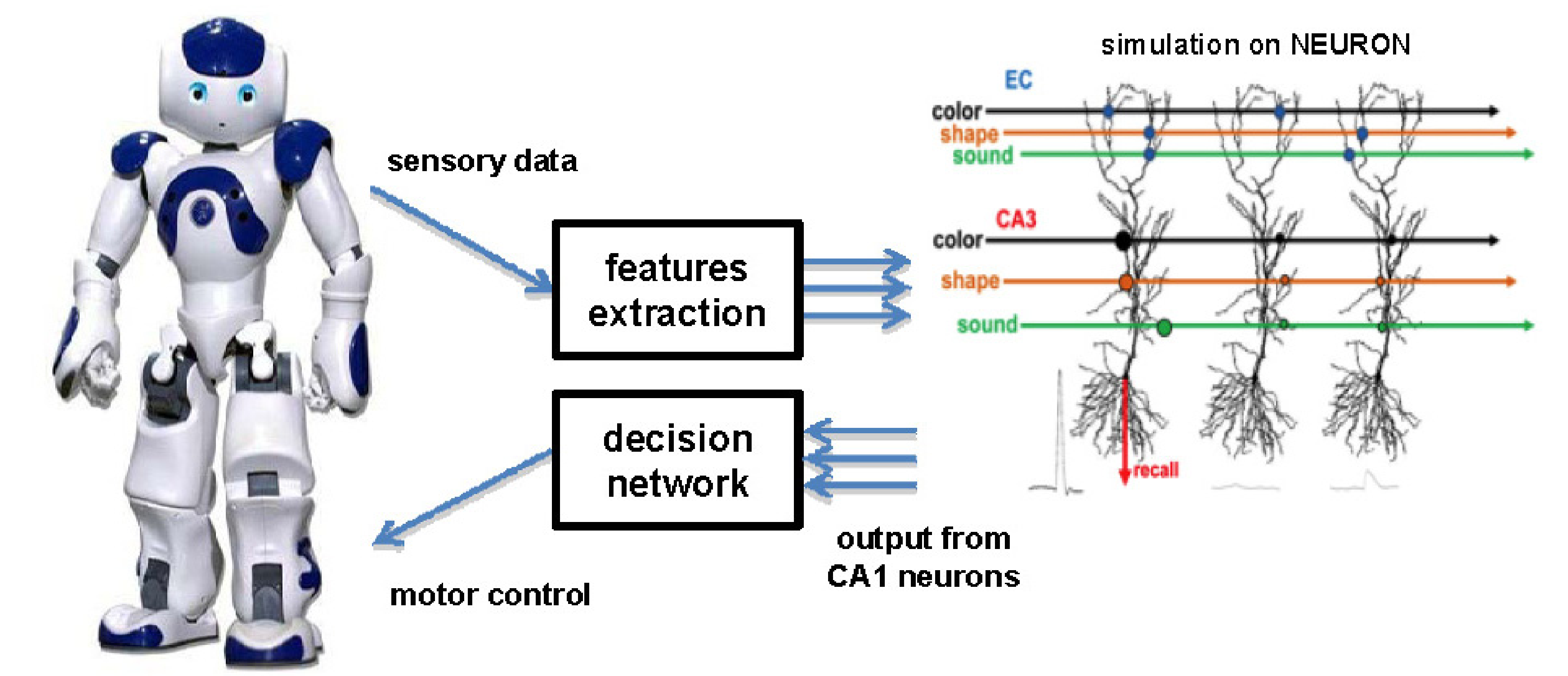

Map
CameraImage
File
View
Obstacle
Capture
Localization
Robot
?
SET ROBOT
Lasers

ragazzi, museo e
Pubblica Amministrazione:
un robot per la cultura

ROBOBICO Project

- Joint research with Michele Migliore, Institute of Biophysics, National Research Council, Palermo, Italy.
- Control a humanoid robot by a realistic network of morphologically accurate neurons.
- To learn about the relationships between structure, dynamics, functions and dysfunctions of neuronal circuits.
- To produce experimentally testable predictions facilitating the development of innovative drugs and therapies.

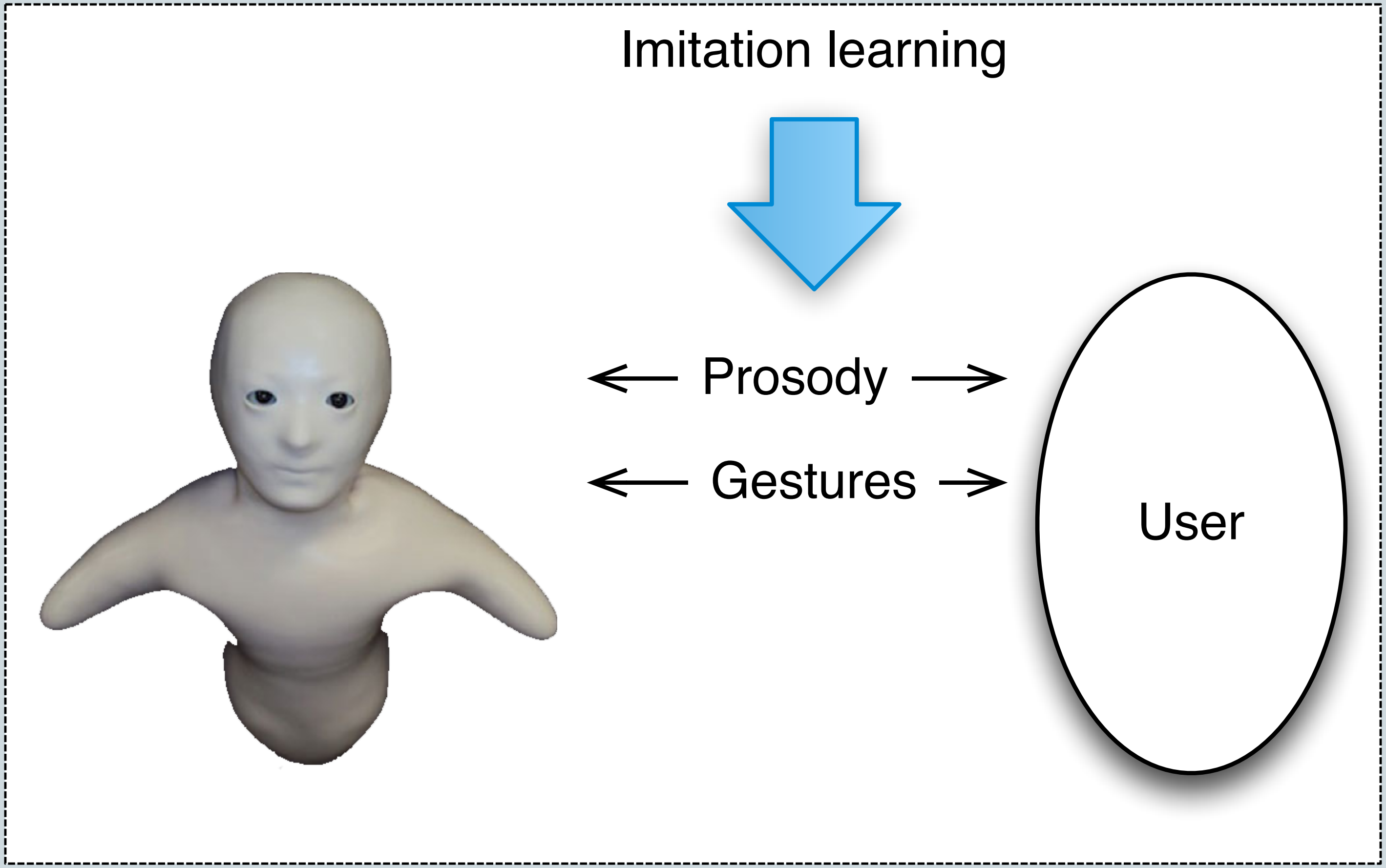
Loop between real-world and neuronal model



Role of Embodied Interaction for MC

- Collaboration with Hiroshi Ishiguro (Osaka University)
- Essential embodiment
- Motions of head, lips, arms
- Sense of presence

Basic Embodiment for Interactions



Imitation learning

← Prosody →

← Gestures →

User

A Slogan for RoboticsLab

BUILDING ARTIFACTS ABLE TO IMPROVE OUR INNER LIFE

Thank you for your attention!

Viale delle Scienze, Palermo, Sicily, Italy
L'indirizzo è approssimativo

antonio.chella@unipa.it