

Assessing and Characterizing the Cognitive Power of Machine Consciousness Implementations

Raúl Arrabales, Agapito Ledezma and Araceli Sanchis

Computer Science Department

Carlos III University of Madrid

<http://Conscious-Robots.com/Raul>

AAAI 2009 Fall Symposium Series
Biologically Inspired Cognitive Architectures 2009
Washington, D.C. November 5-7.



**CONSCIOUS
ROBOTS**

www.Conscious-Robots.com

➔ Contents

- Context and Objectives
- Related Work
- *ConsScale*
- Applying the scale
- Conclusions

→ Context

- **Evaluation** of cognitive architectures.
 - To help to decide what design strategies are better.
 - To determine the state of the art.
 - To enable comparative analysis.
- **Characterization** of cognitive architectures.
 - Strengths and weaknesses.
 - Key features and lacks.

➔ Objectives

- **Objectives** of this work:

- To establish a **scale** for assessing the cognitive power of artificial architectures.
- To define a generic framework (problem domain independent).

- **Current Limitations:**

- Specifically oriented to *Machine Consciousness*.
- Inspired on natural phylogeny and ontogeny.
- Not specifically aimed at Phenomenal Consciousness.

➔ Related Work

■ **Consciousness Correlates and Metrics.**

- Clinical diagnosis:
 - GCS, JFK-revised, ERP, Bispectral I., Gamma synch., etc.
- Information integration measures:
 - Φ (Tononi), Causal Density (Seth), etc.
- Hallmarks, axiomatic properties, and constraints:
 - Arch. (Sloman), Axioms (Aleksander), Constr. (Metzinger).
- Potential correlates of consciousness:
 - OMC (Gamez).
- Behavioral correlates of consciousness:
 - T. Test (Turing), Adapted TT (Harnad), reportability, etc.

➔ ConsScale

- Combines third-person approaches with inner inspection.
- *ConsScale* Criteria:
 - Cognitive skills and associated observed behavior.
 - Definition of a set of **generic cognitive skills**.
 - Architecture inspection and component analysis.
 - Definition of a set of **abstract architectural components**.

➔ Architectural Components

E	Environment. World around the agent.
B	Body. Physical or simulated agent body.
S_{ext}	Exteroceptive sensors.
S_{proprio}	Proprioceptive sensors.
A	Action Machinery. Agent effectors.
R	Reasoning. Sensorimotor coordination machinery.
M	Memory. Internal agent state.
Att	Attention mechanism to direct S and A to a specific E_i
Mⁿ	Capability of multiple context representation.
SsA	Self-status assessment mechanism.
I	Mechanism for the representation of the self.
O	Mechanism for the representation of other selves.
AR	Accurate report mechanism.
AVR	Accurate verbal report mechanism.
R_n	Mechanism to run and synchronize several streams of consciousness.

Cognitive Skills ($CS_{i,j}$)

Level	Cognitive Skills ($CS_{i,j}$)
2	$CS_{2,1}$: Fixed reactive responses ("reflexes").
3	$CS_{3,1}$: Autonomous acquisition of new adaptive reactive responses. $CS_{3,2}$: Usage of proprioceptive sensing for embodied adaptive responses.
4	$CS_{4,1}$: Selection of relevant sensory information. $CS_{4,2}$: Selection of relevant motor information. $CS_{4,3}$: Selection of relevant memory information. $CS_{4,4}$: Evaluation (positive or negative) of selected objects or events. $CS_{4,5}$: Selection of what needs to be stored in memory. $CS_{4,6}$: Trial and error learning. Re-evaluation of selected objects or events. $CS_{4,7}$: Directed behavior toward specific targets like following or escape. $CS_{4,8}$: Evaluation of the performance in the achievement of a single goal. $CS_{4,9}$: Basic planning capability: calculation of next n sequential actions. $CS_{4,10}$: Depictive representations of percepts [17].
5	$CS_{5,1}$: Ability to move back and forth between multiple tasks. $CS_{5,2}$: Seeking of multiple goals. $CS_{5,3}$: Evaluation of the performance in the achievement of multiple goals. $CS_{5,4}$: Autonomous reinforcement learning (emotional learning). $CS_{5,5}$: Advanced planning capability considering all active goals.
6	$CS_{6,1}$: Self-status assessment (background emotions). $CS_{6,2}$: Background emotions cause effects in agent's body. $CS_{6,3}$: Representation of the effect of emotions in organism (feelings). $CS_{6,4}$: Ability to hold a precise and updated map of body schema. $CS_{6,5}$: Abstract learning (learned lessons generalization).

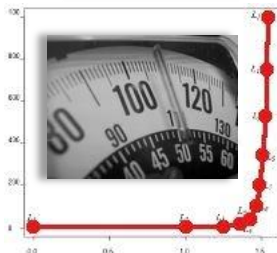
7	$CS_{7,1}$: Representation of the relation between self and perception. $CS_{7,2}$: Representation of the relation between self and action. $CS_{7,3}$: Representation of the relation between self and feelings. $CS_{7,4}$: Self-recognition capability. $CS_{7,5}$: Advance planning including the self as an actor in the plans. $CS_{7,6}$: Use of imaginational states in planning. $CS_{7,7}$: Learning of tool usage.
8	$CS_{8,1}$: Ability to model others as subjective selves. $CS_{8,2}$: Learning by imitation of a counterpart. $CS_{8,3}$: Ability to collaborate with others in the pursuit of a common goal. $CS_{8,4}$: Social planning (planning with socially aware plans). $CS_{8,5}$: Ability to make new tools.
9	$CS_{9,1}$: Ability to develop Machiavellian strategies like lying and cunning. $CS_{9,2}$: Social learning (learning of new Machiavellian strategies). $CS_{9,3}$: Advanced communication skills (accurate report of mental content). $CS_{9,4}$: Groups are able to develop a culture.
10	$CS_{10,1}$: Accurate verbal report. Advanced linguistic capabilities. $CS_{10,2}$: Ability to pass the Turing test. $CS_{10,3}$: Ability to modify and adapt the environment to agent's needs. $CS_{10,4}$: Groups are able to develop a civilization and advance culture and technology.
11	$CS_{11,1}$: Ability to manage several streams of consciousness.

➔ *ConsScale* features

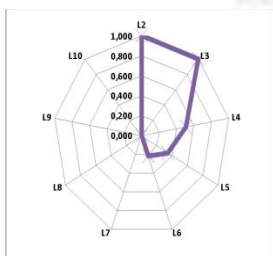
- Characterization is achieved using these tools:



- Definition of an ordered **list of levels** associated with consciousness (from -1 to 11).

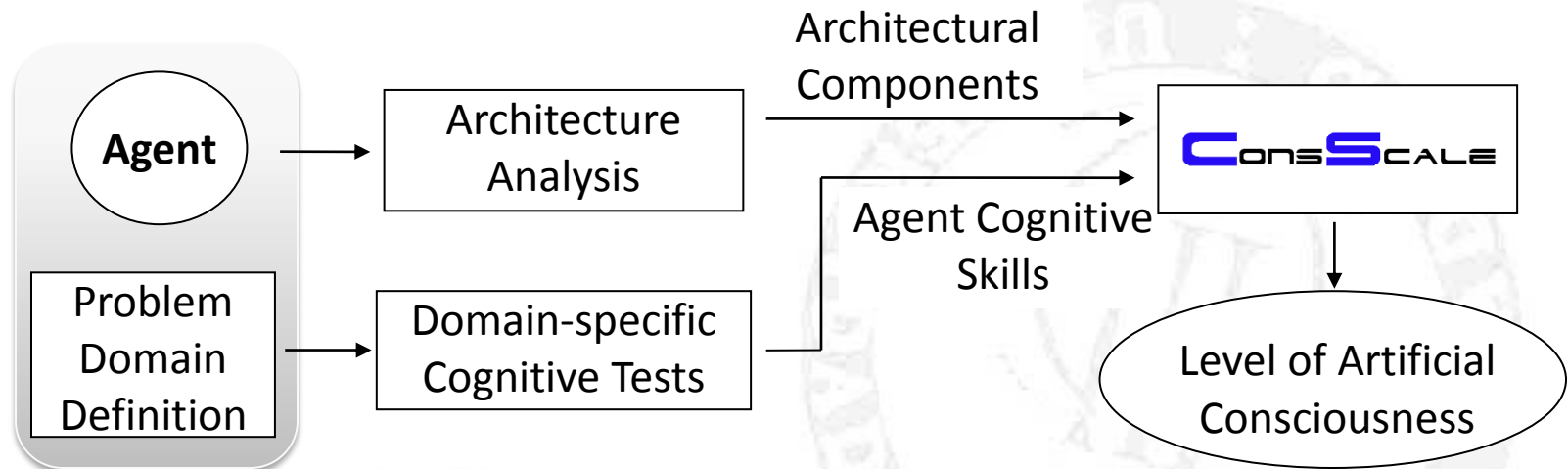


- Definition of a **Quantitative Score** (CQS: from 0 to 1000).

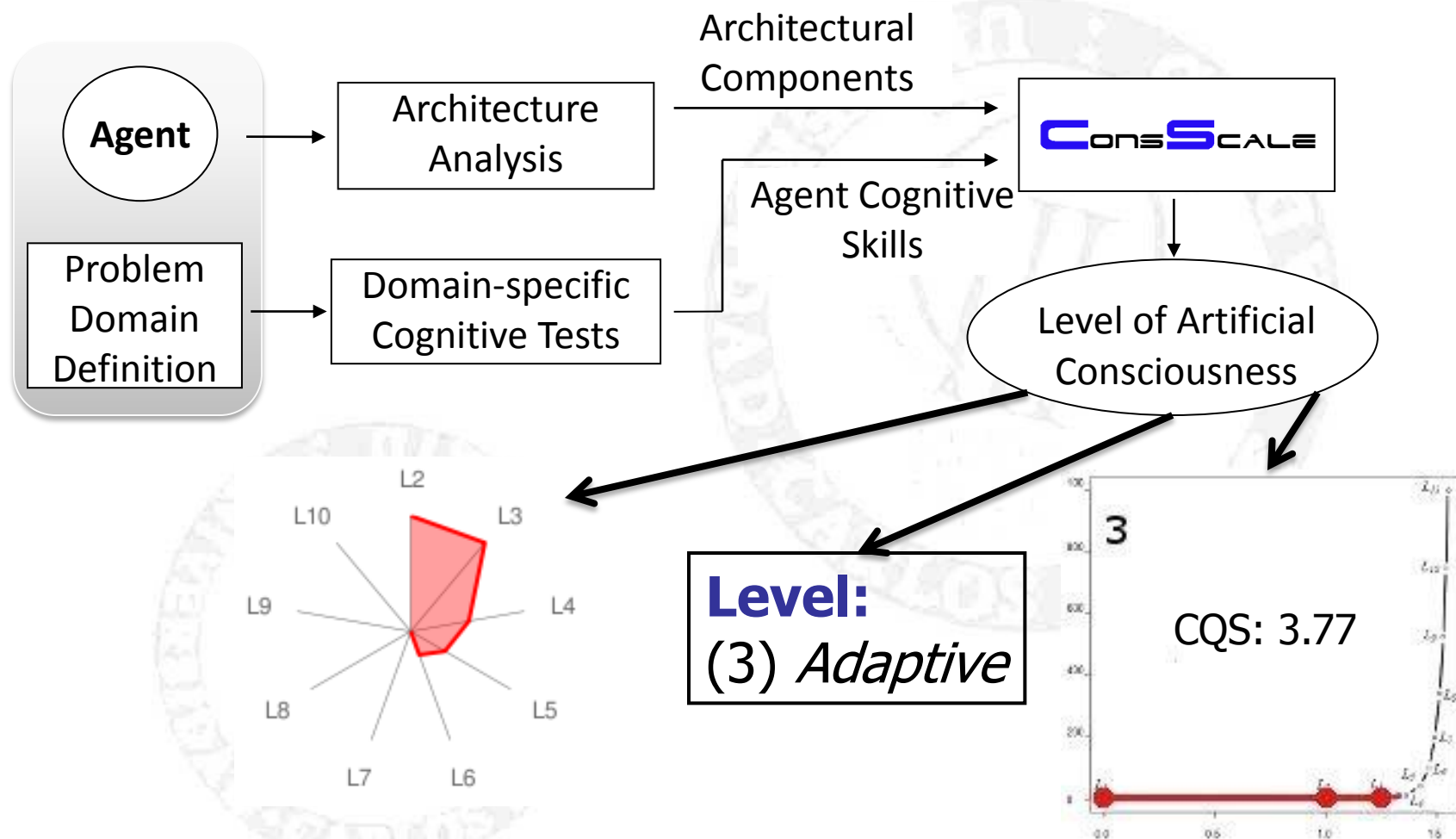


- Definition of a graphical **cognitive profile** (from L2 to L11).

➔ *ConsScale* evaluation process



➔ *ConsScale* evaluation process



➔ *ConsScale* Evaluation Steps

- Select a cognitive architecture implementation.
- Identify instantiated architectural components.
- Select problem domain.
- For each $CS_{i,j}$ define domain-specific cognitive test/s.
- *ConsScale* Assessment:
 - Assign *ConsScale* Level.
 - Calculate *ConsScale* Quantitative Score.
 - Generate *ConsScale* Graphical Cognitive Profile.

➔ ConsScale Levels

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

0.00

0.18

2.22

12.21

41.23

101.08

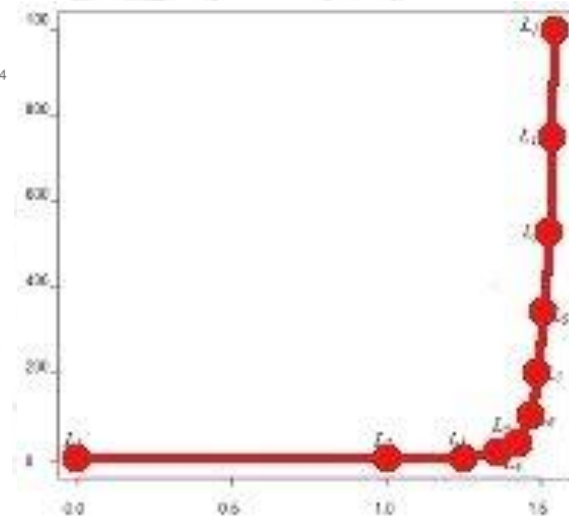
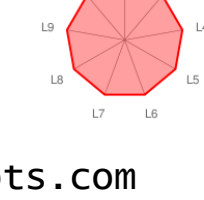
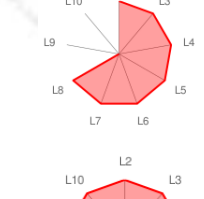
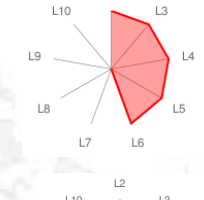
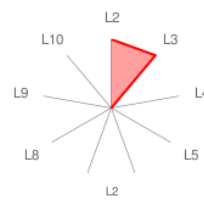
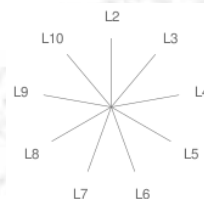
200.03

341.45

524.54

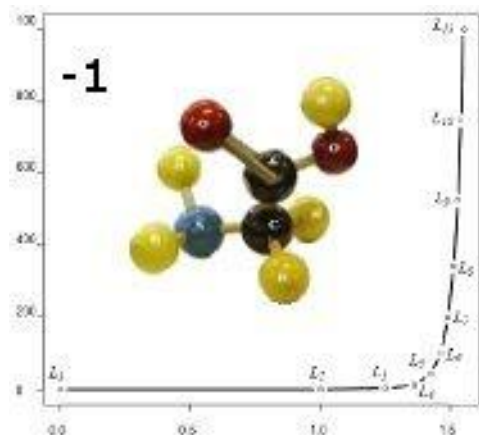
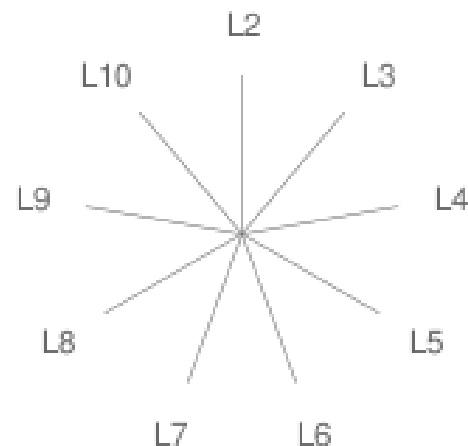
745.74

1000.00



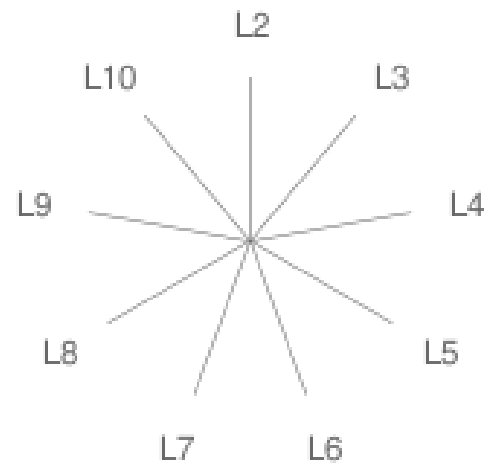
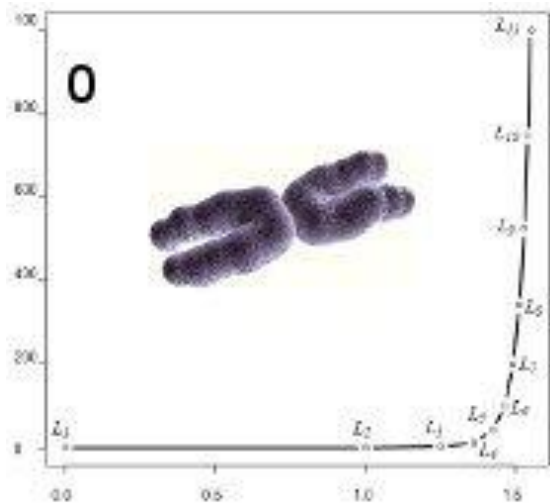
➔ *ConsScale* (I)

- Level -1: *Disembodied*
 - Behavior: not a situated agent.
 - Phylogeny: amino acid

**E****B**

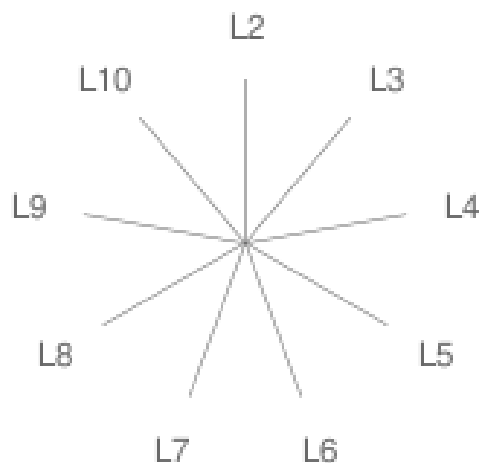
➔ *ConsScale* (II)

- Level 0: *Isolated*
 - Behavior: not a situated agent.
 - Phylogeny: isolated chromosome.

E**B**

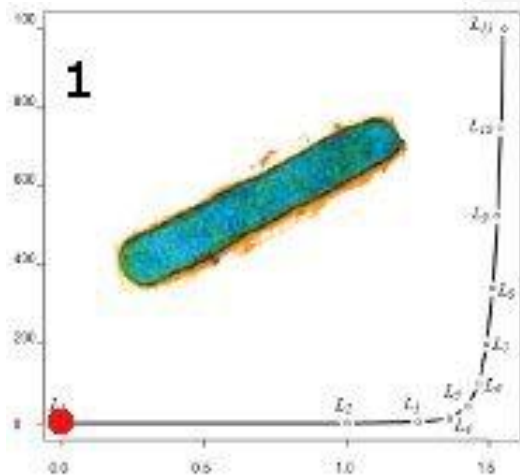
➔ *ConsScale* (III)

- Level 1: *Decontrolled*
 - Behavior: not a situated agent.
 - Phylogeny: dead bacteria.



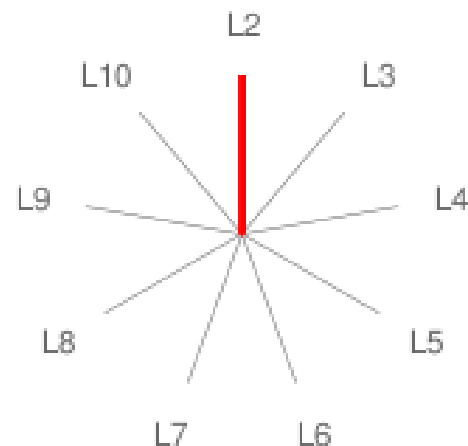
E

S B A



➔ *ConsScale* (IV)

- Level 2: *Reactive*
 - Behavior: reflexes.
 - Phylogeny: virus.



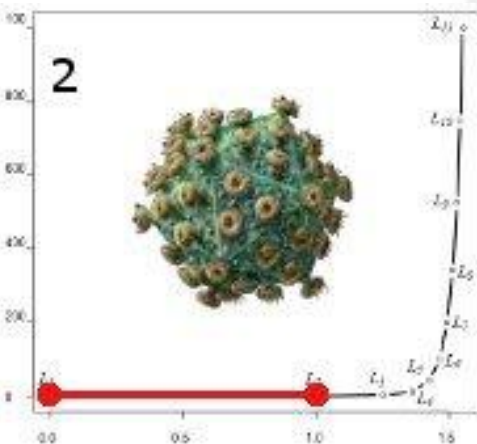
E

B

S

R

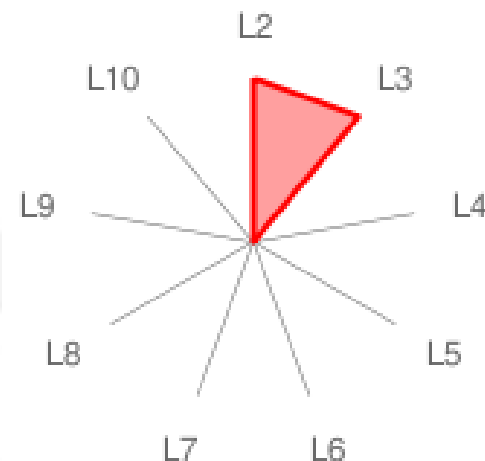
A



➔ *ConsScale (V)*

■ Level 3: *Adaptive*

- Behavior: basic ability to learn new reflexes.
- Phylogeny: earthworm.



E

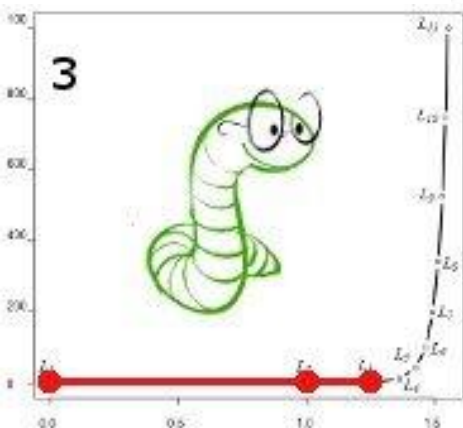
B

M

S

R

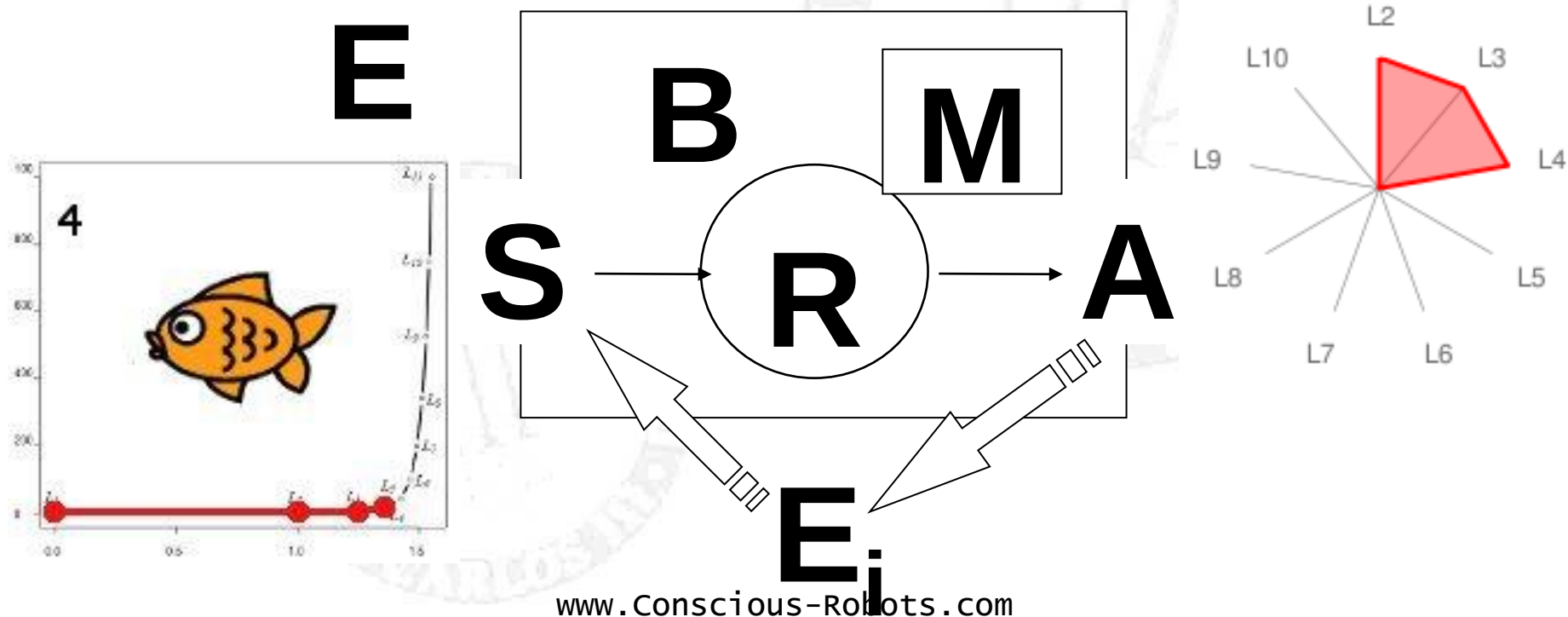
A



➔ *ConsScale* (VI)

■ Level 4: *Attentional*

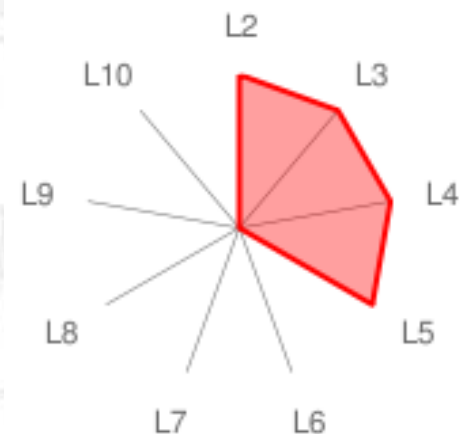
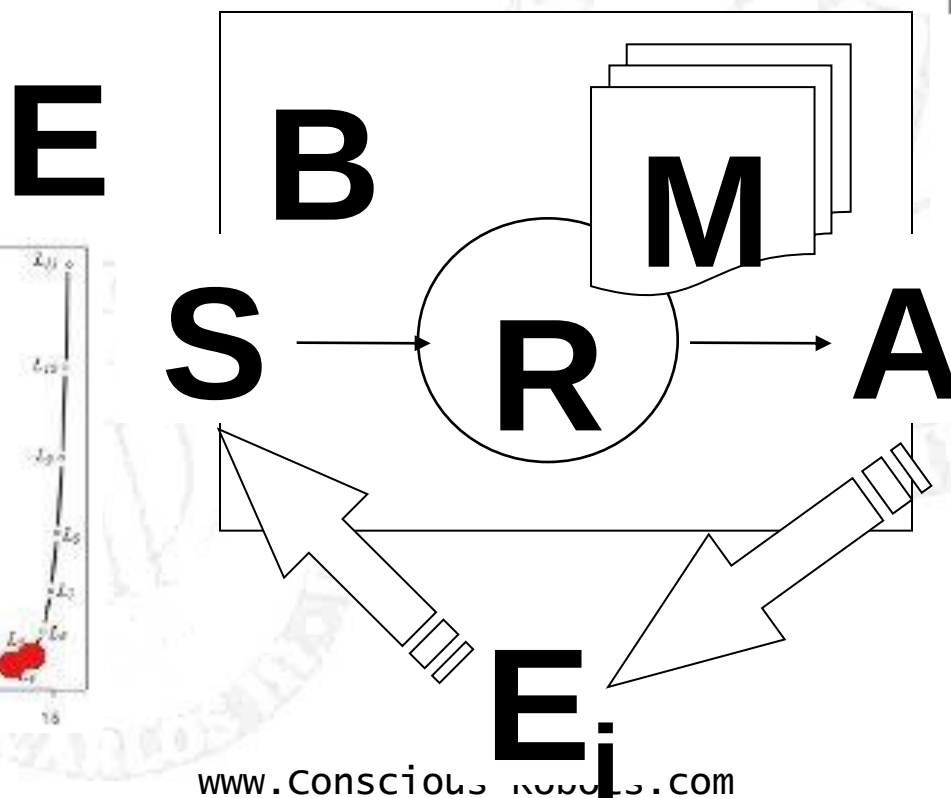
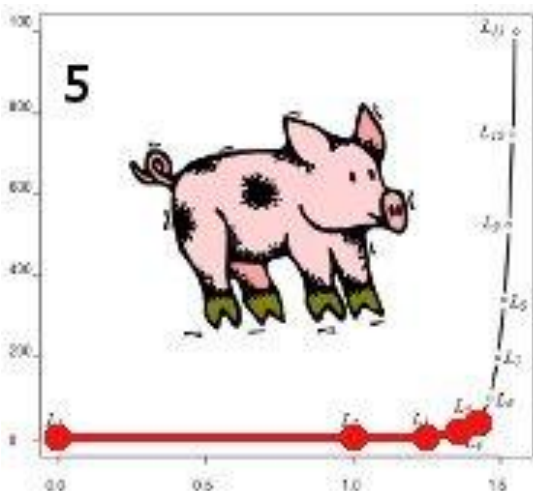
- Behavior: attack and escape. Attention + emotion.
- Phylogeny: fish.



➔ *ConsScale* (VII)

■ Level 5: *Executive*

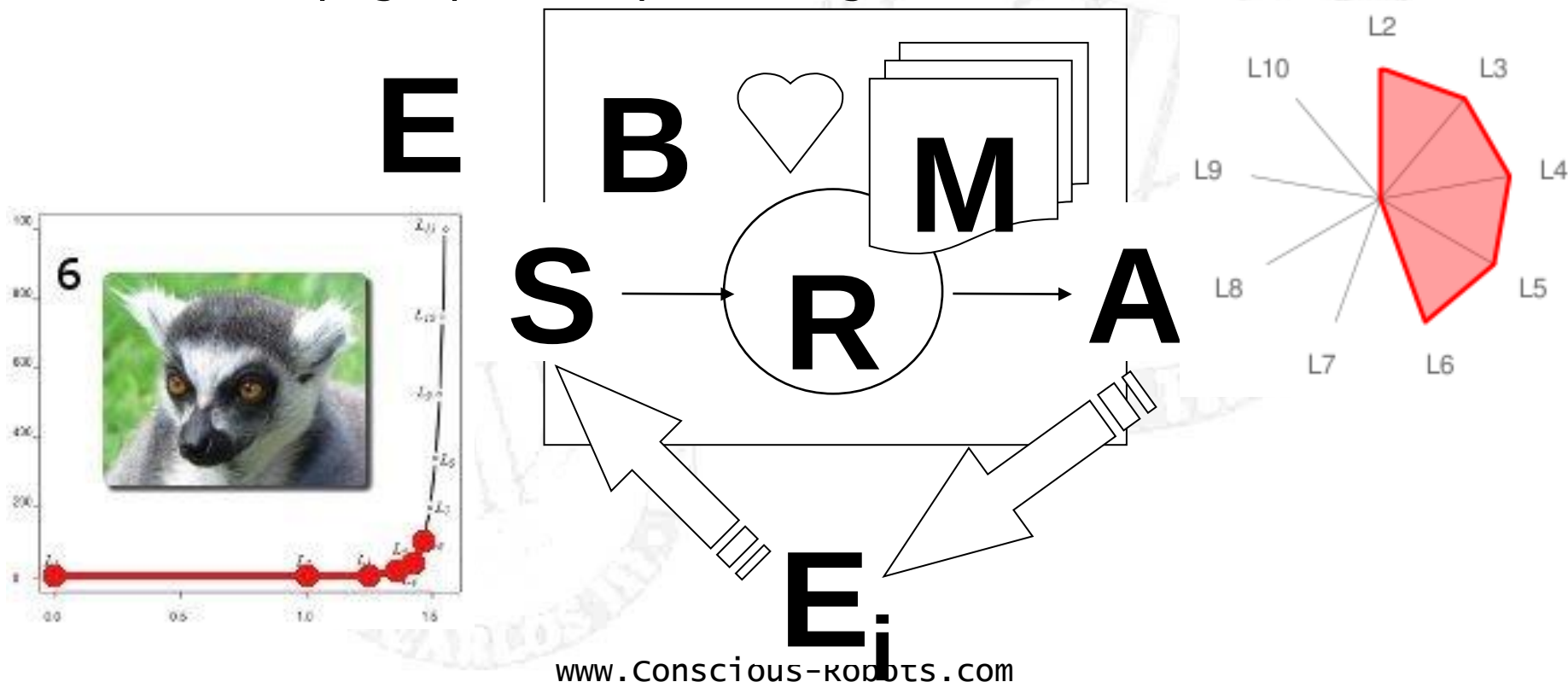
- Behavior: set shifting. Emotional learning.
- Phylogeny: quadruped mammal.



➔ *ConsScale* (VIII)

■ Level 6: *Emotional*

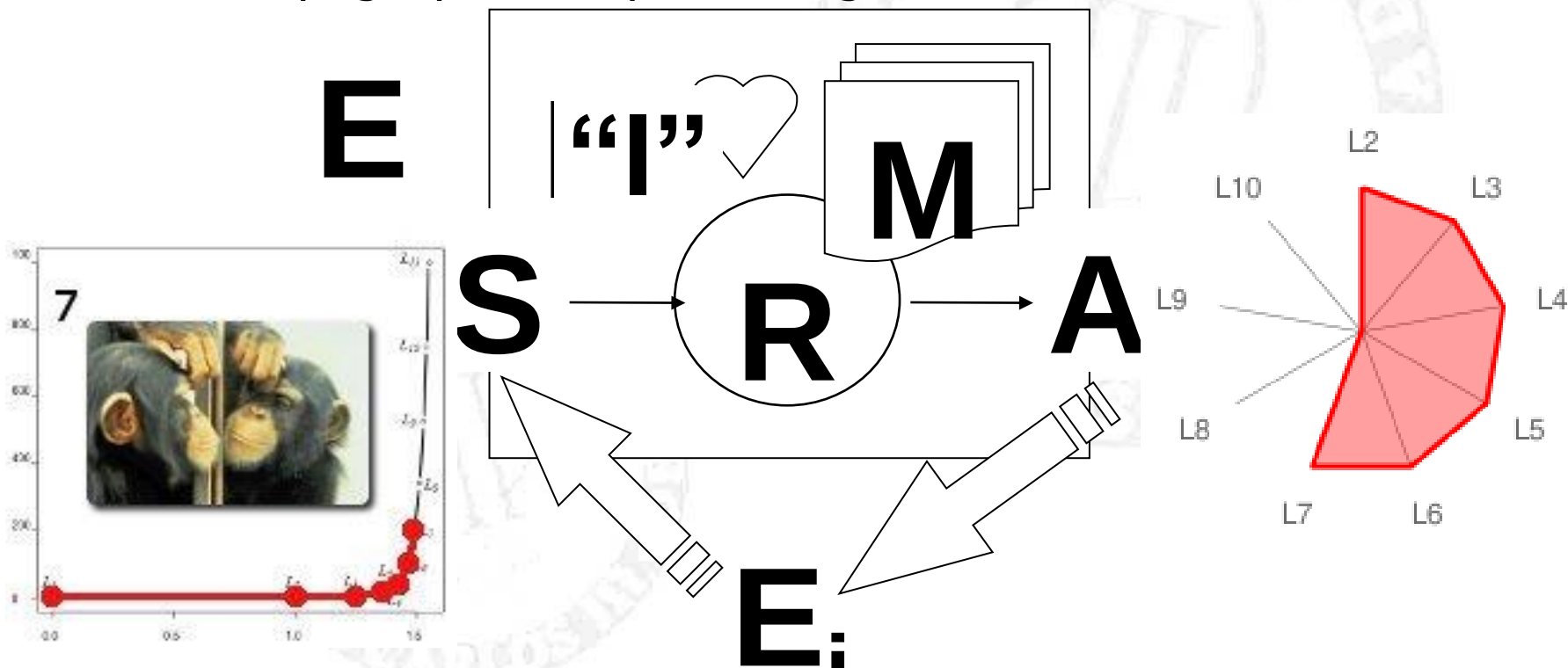
- Behavior: feelings influence behavior.
- Phylogeny: monkey. ToM Stage 1: "I know".



➔ *ConsScale* (IX)

■ Level 7: *Self-Conscious*

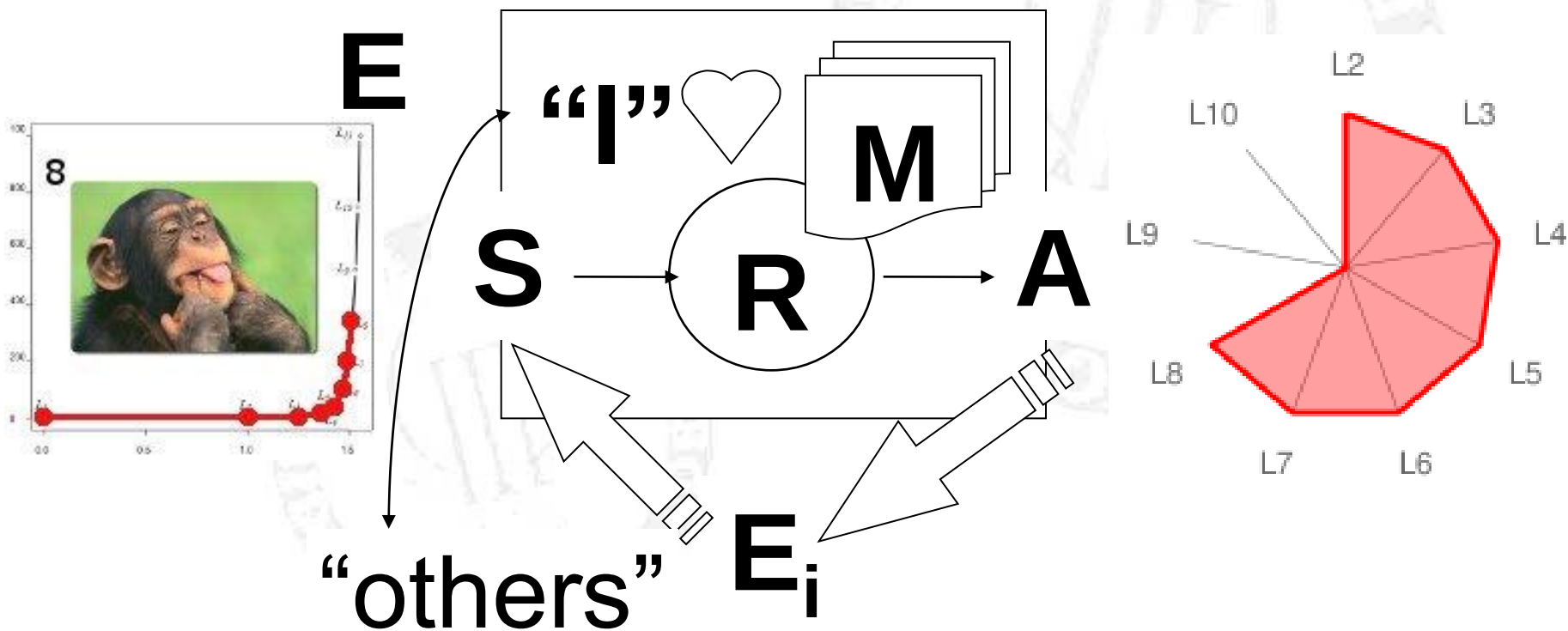
- Behavior: advanced planning. Usage of tools. Mirror test.
- Phylogeny: monkey. ToM Stage 2: "I know I know".



➔ *ConsScale (X)*

■ Level 8: *Empathic*

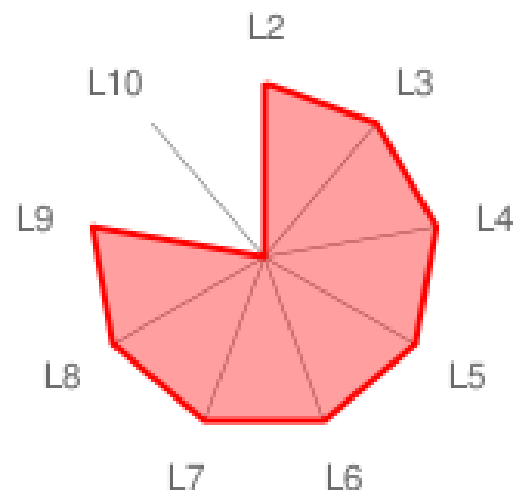
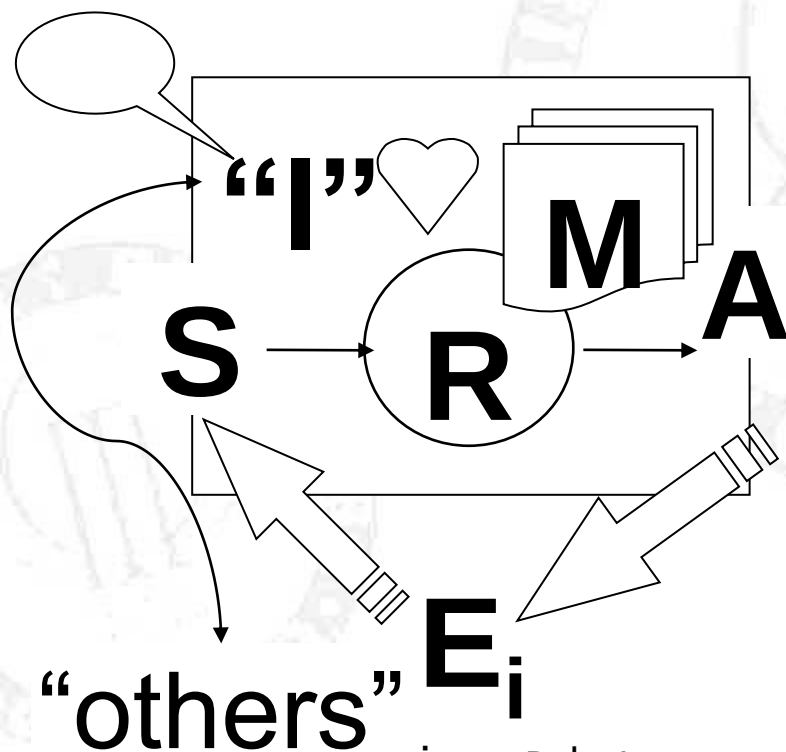
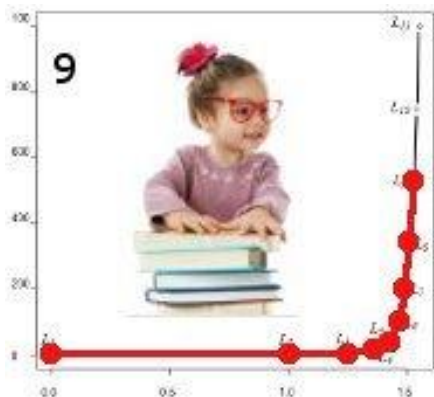
- Behavior: making of tools. Social behavior.
- Phylogeny: chimpanzee. ToM Stage 3: "I know you know".



➔ *ConsScale* (XI)

■ Level 9: *Social*

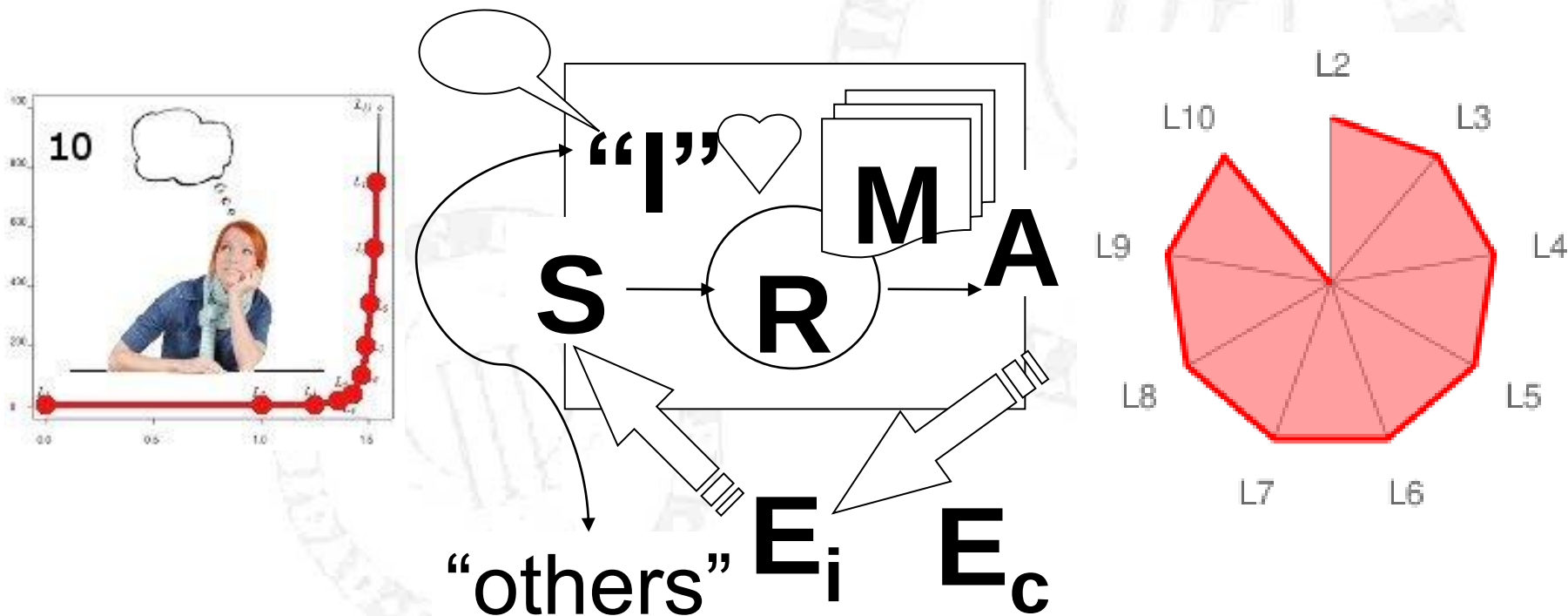
- Behavior: linguistic capabilities. Ability for culture.
- Phylogeny: human. ToM Stage 4: "I know you know I know".



➔ *ConsScale* (XII)

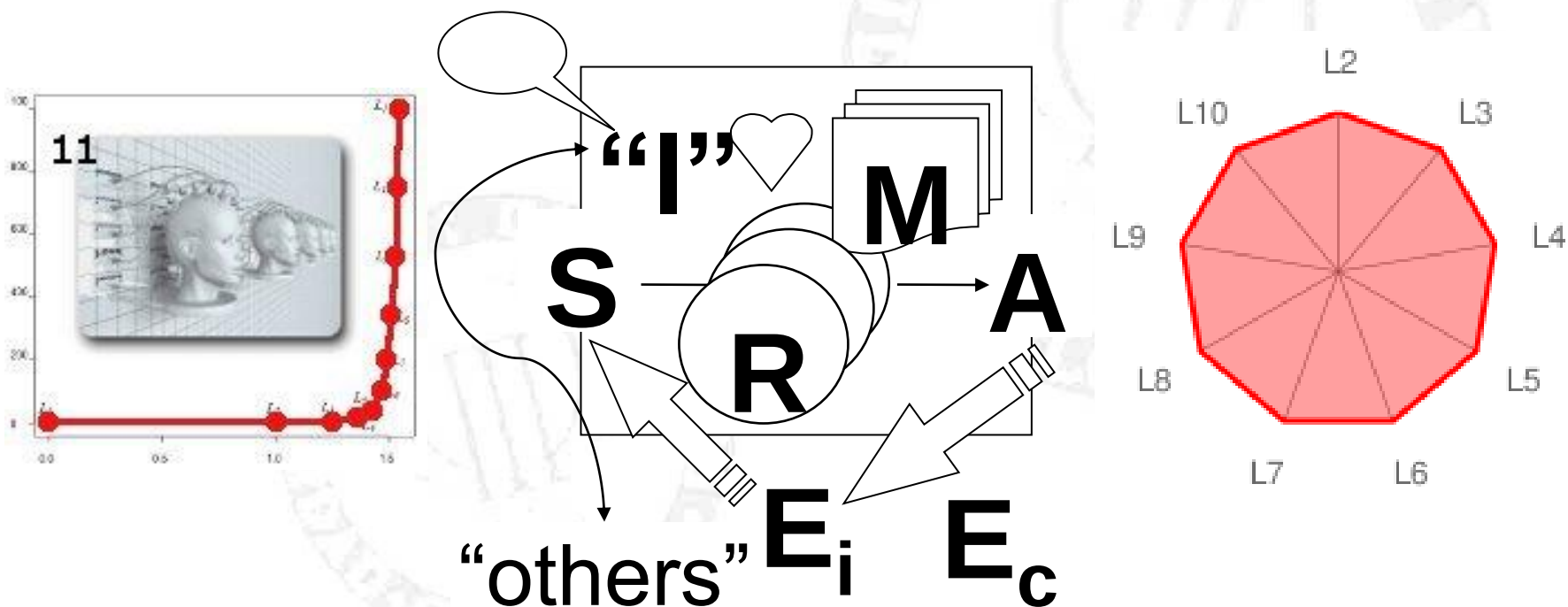
■ Level 10: *Human-like*

- Behavior: accurate verbal report. Culture. Technology.
- Phylogeny: human. Adapted environment.



➔ *ConsScale* (XIII)

- Level 11: *Super-Conscious*
 - Behavior: several streams of consciousness.
 - Phylogeny: n/a.



➔ *ConsScale* Characterization

■ Agent L3

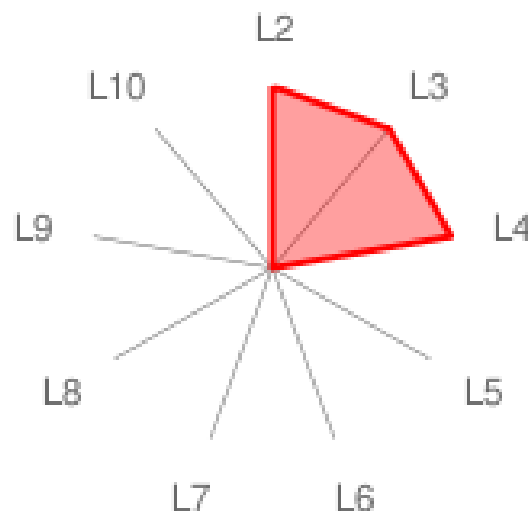
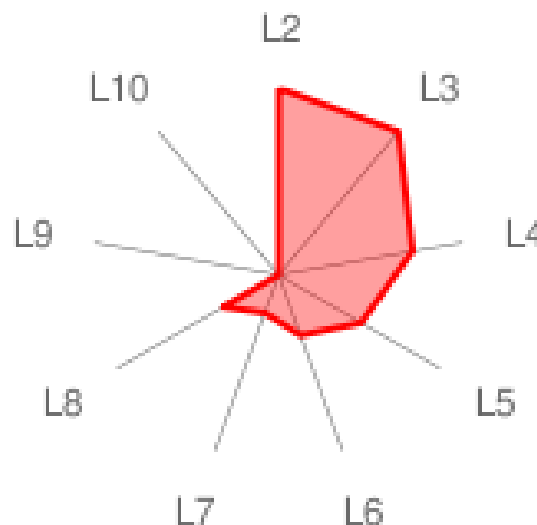
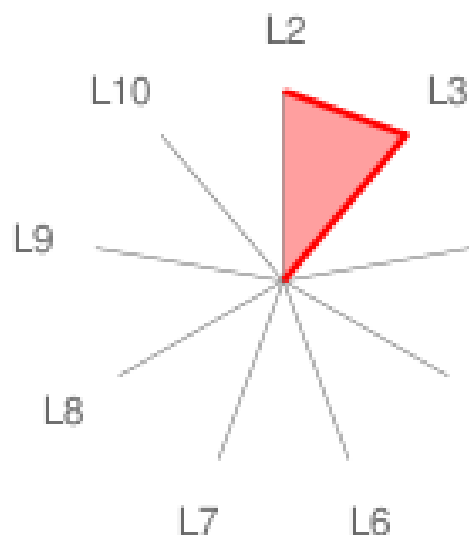
- Level: 3 - *Adaptive* (canonical)
- CQS: 2.22

■ Agent L3+

- Level: 3 – *Adaptive* (uncanonical)
- CQS: 9.04

■ Agent L4

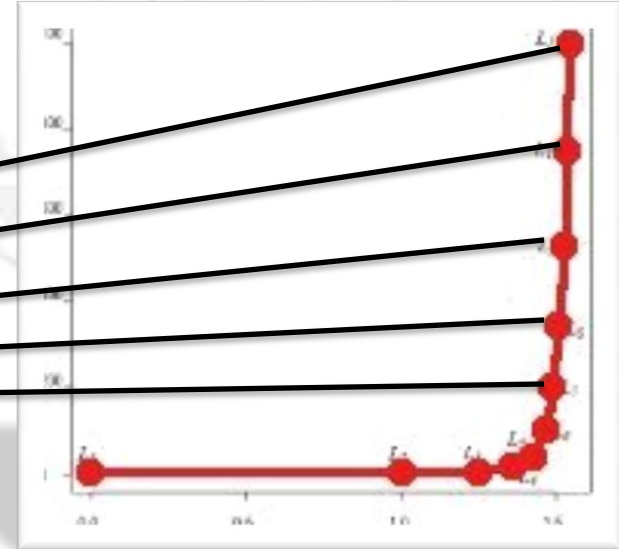
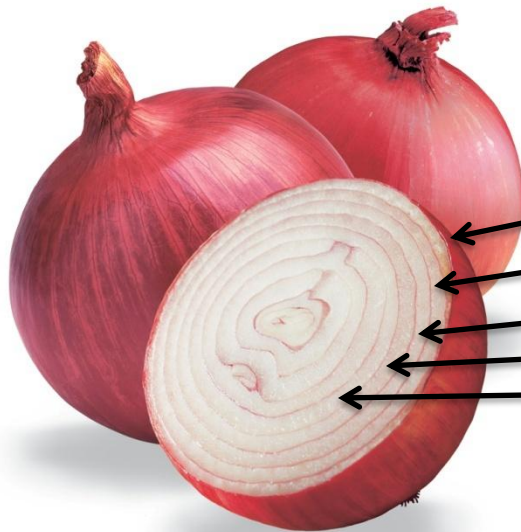
- Level: 4 - *Attentional* (canonical)
- CQS: 12.21



Agents qualify as level n ***if and only if*** all lower levels are also satisfied.

➔ *ConsScale* as a Roadmap

- Considers consciousness as an integrator that puts a mind together.
- Considers **synergy** instead of isolated cognitive abilities.
- Evolutionary inspired developmental path.



→ CQS

- *ConsScale* Quantitative Score.

- L_i

- Particular score for level i .

- CLS

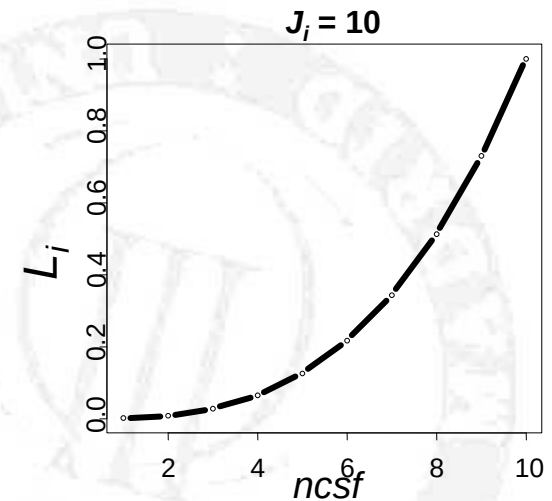
- Cumulative Level score.

- CQS

- ConsScale Quantitative score.

➔ L_i

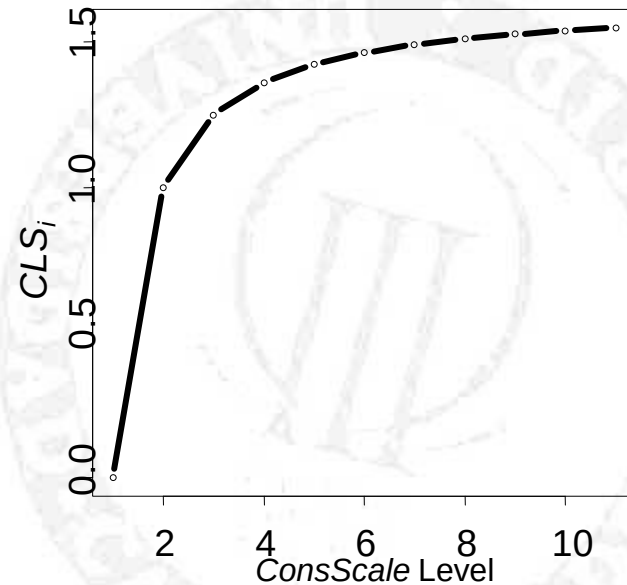
$$L_i = \begin{cases} 0 & \text{if } ncsf \text{ is } 0 \\ \frac{ncsf + (1 - J_i)}{10^3} & \text{otherwise} \end{cases}$$



- $ncsf \rightarrow$ Number of cognitive skills fulfilled.
- $J \rightarrow$ Maximum number of CS.
- $J_i \rightarrow$ Total number of CS in level i.

➔ *CLS*

$$CLS = \sum_{i=2}^{11} \left(\frac{L_i}{i-1} \right)^2$$



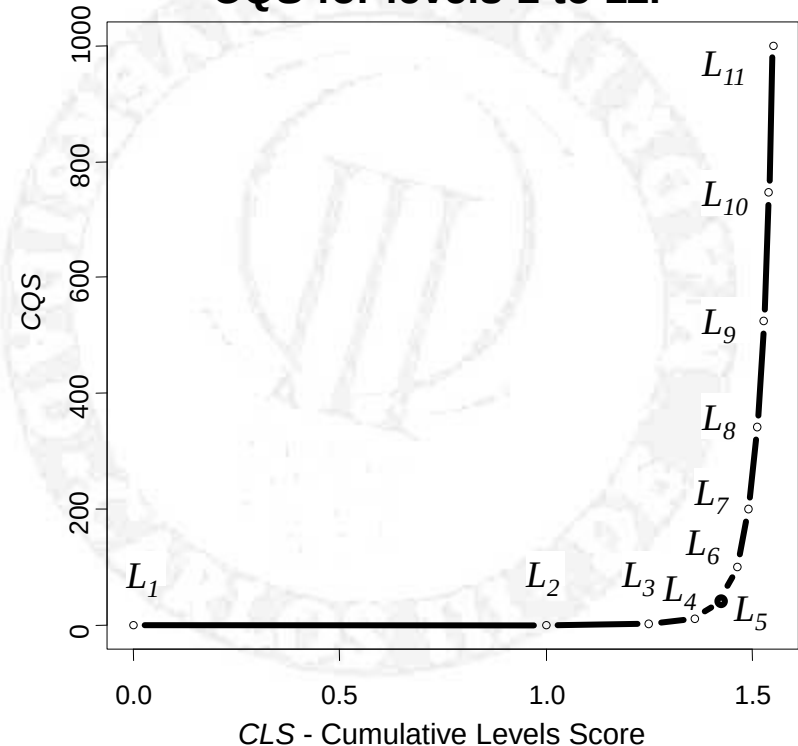
- $L_i \rightarrow$ Level i partial score.
- $i \rightarrow$ Level index.

➔ CQS (0 to 1000)

$$CQS = \frac{e^{(CLS^5/K)} + a}{10}$$

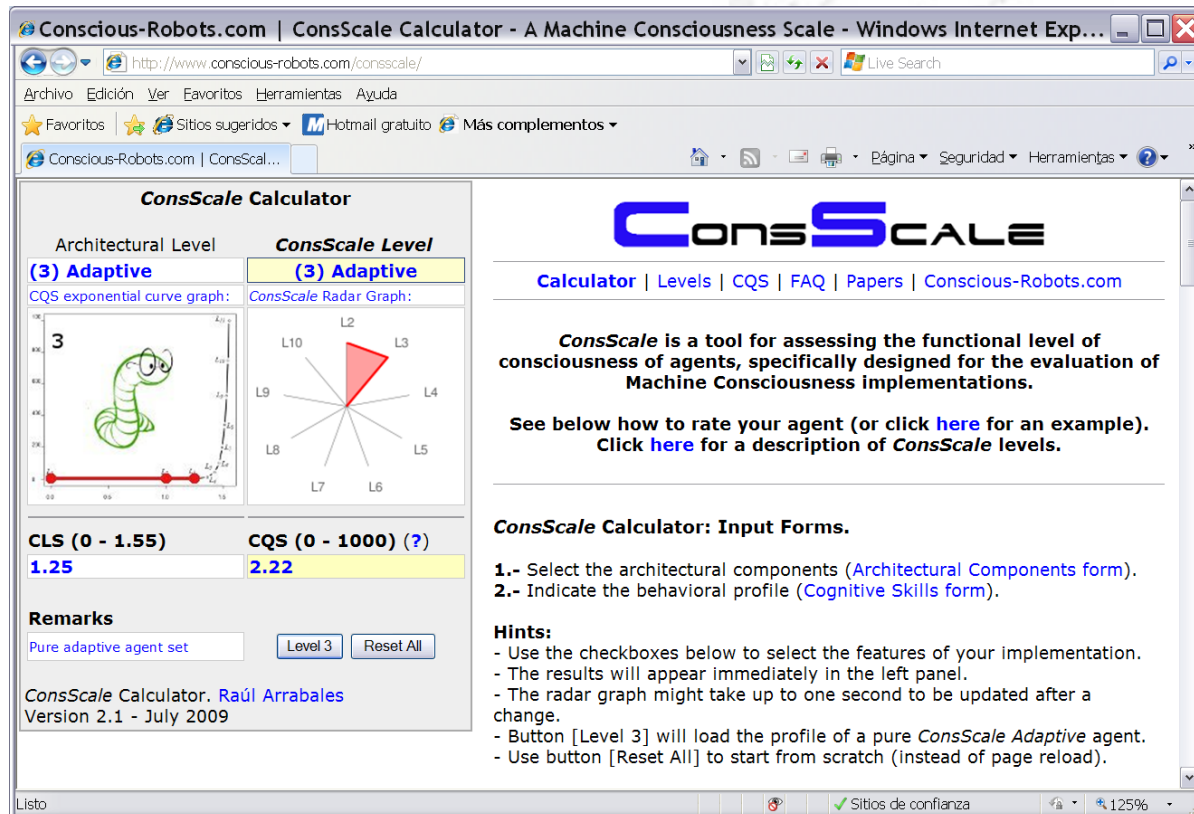
- $CLS \rightarrow$ Cumulative level score.
- $K \rightarrow$ Constant value ($\sim 0'97$).
- $a \rightarrow$ Constant value (-1).

CQS for levels 1 to 11.



➔ CQS online calculator

<http://conscious-robots.com/consscale>



The screenshot shows the ConsScale Calculator web application in a Windows Internet Explorer browser window. The page title is "ConsScale Calculator - A Machine Consciousness Scale". The browser address bar shows the URL "http://www.conscious-robots.com/consscale/".

The application interface is divided into two main sections. The left section, titled "ConsScale Calculator", contains the following elements:

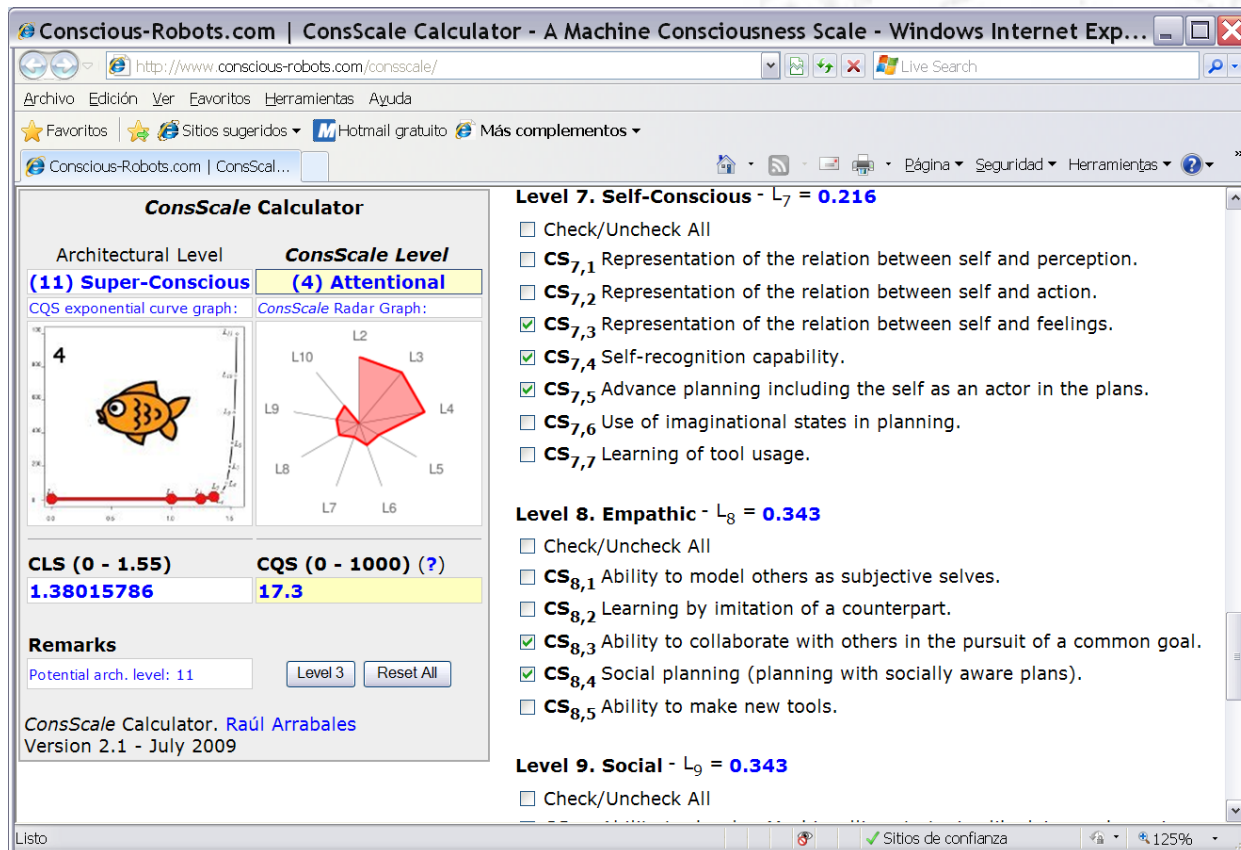
- Architctural Level**: (3) Adaptive
- ConsScale Level**: (3) Adaptive
- CQS exponential curve graph**: A graph showing a green curve with a value of 3.
- ConsScale Radar Graph**: A radar graph with 10 axes (L1 to L10) and a red shaded area.
- CLS (0 - 1.55)**: 1.25
- CQS (0 - 1000) (?)**: 2.22
- Remarks**: Pure adaptive agent set
- Buttons**: Level 3, Reset All
- Footer**: ConsScale Calculator. Raúl Arrabales, Version 2.1 - July 2009

The right section, titled "ConsScale", contains the following elements:

- ConsScale Logo**
- Navigation Links**: Calculator | Levels | CQS | FAQ | Papers | Conscious-Robots.com
- Description**: ConsScale is a tool for assessing the functional level of consciousness of agents, specifically designed for the evaluation of Machine Consciousness implementations.
- Instructions**: See below how to rate your agent (or click here for an example). Click here for a description of ConsScale levels.
- ConsScale Calculator: Input Forms.**
 - 1.- Select the architectural components (Architctural Components form).
 - 2.- Indicate the behavioral profile (Cognitive Skills form).
- Hints:**
 - Use the checkboxes below to select the features of your implementation.
 - The results will appear immediately in the left panel.
 - The radar graph might take up to one second to be updated after a change.
 - Button [Level 3] will load the profile of a pure ConsScale Adaptive agent.
 - Use button [Reset All] to start from scratch (instead of page reload).

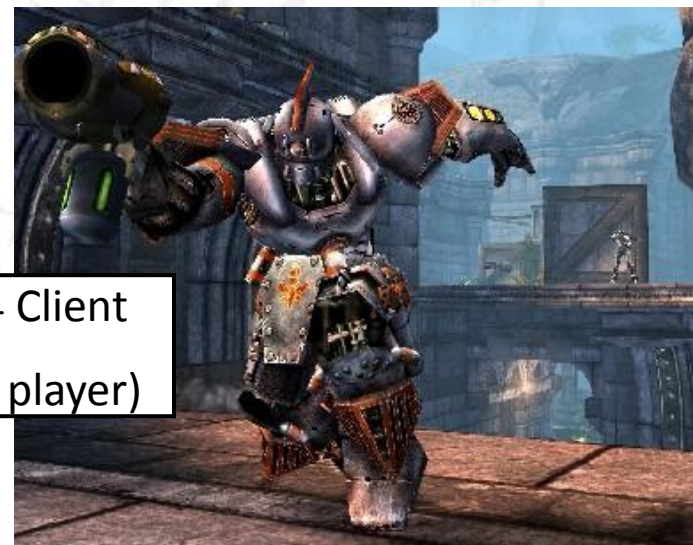
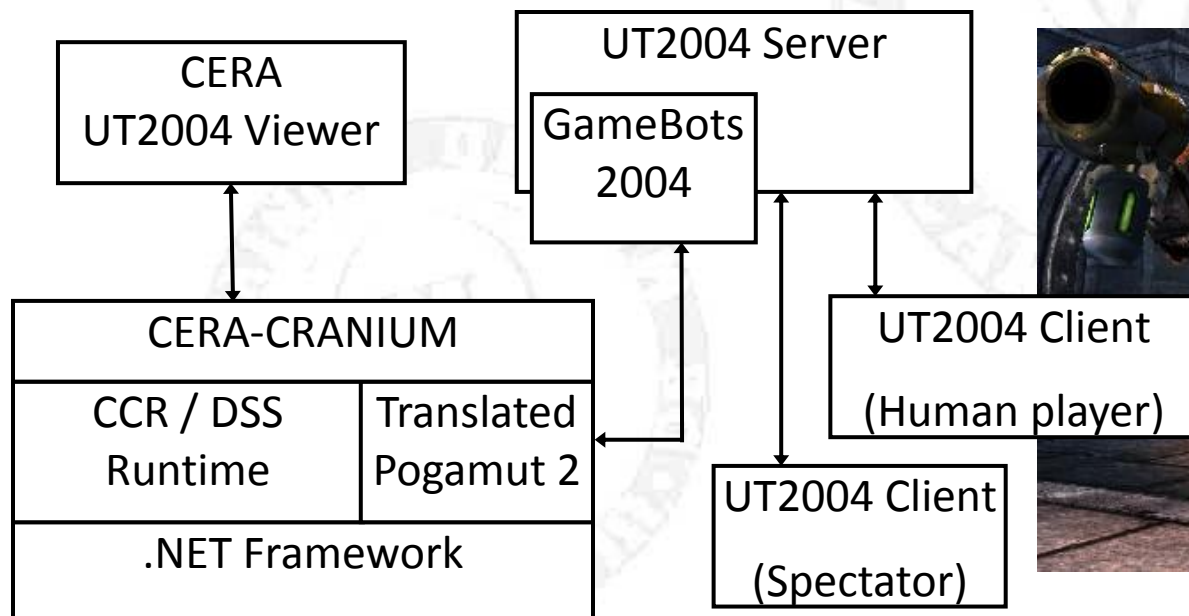
➔ CQS online calculator

<http://conscious-robots.com/consscale>



➔ Applying the scale

- Generic levels have to be instantiated:
 - Generic Cognitive Skills → Specific Behavioral Profiles**
 - Behavioral Profiles → Domain-Specific Behavioral Tests**
- Example: First-Person Shooter Video Game Autonomous Bot.



➔ The Video Game Bot Domain

- Formal definition of the problem domain:
 - **Percepts:** what the sensors of the agent can acquire.
 - **Actions:** what the effectors of the agent can do.
- Instantiation process consists of re-defining $CS_{i,j}$ using a domain-specific ontology:

Percepts or objects:

Players, Weapons, Ammo.

Actions or verbs:

Move, Jump, Run, Turn, Damage, Fire, Chat.

Note: this ontology doesn't necessarily match the internal representation of the agent.

➔ Behavioral Profiles

- **CS_{3,2}**: Usage of proprioceptive sensing for embodied adaptive responses.
- **BP_{3,2}**: Looking for health packs when health level is low or looking for ammo when needed.





➡ Behavioral Profiles

- **$CS_{4,1}$** : Selection of relevant sensory information.
- **$BP_{4,1}$** : Ignoring detected ammo reloading kits when involved in a firefight and no more ammo is needed.



➔ Behavioral Profiles

- **CS_{4,7}**: Directed behavior toward specific targets like following or escape.
- **BP_{4,7}**: The bot shows directed and sustained behavior towards enemies, like following and shooting them or running away from them.



➔ Behavioral Profiles

- **CS_{4,9}**: Basic ability to plan next movements.
- **BP_{4,9}**: Bot shows a coherent sequence of actions planned in order to reach certain goal. For instance, leaving a firefight for re-arming and then going back to combat.
- **CS_{5,5}**: Ability to plan actions taking into account all active game goals.
- **BP_{5,5}**: Actions are effectively interleaved as required for the accomplishment of multiple active goals. For instance, trajectory is slightly modified whilst chasing and shooting an enemy in order to pick up some ammo packs available in the surroundings.

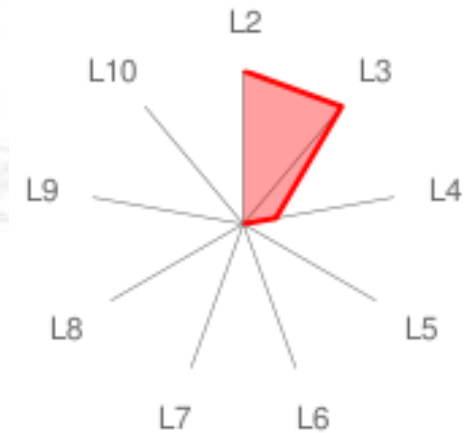
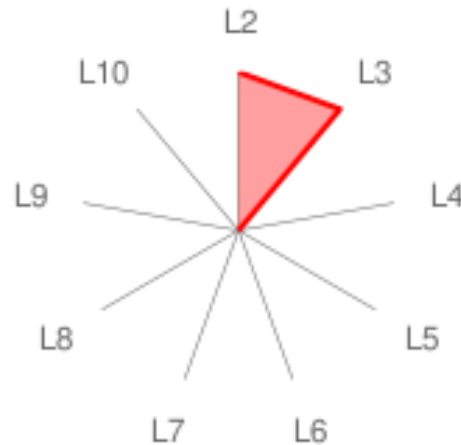
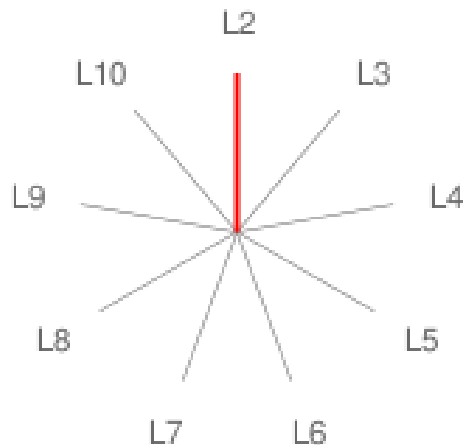
➔ Behavioral Profiles

- **CS_{9,1}**: Ability to develop Machiavellian strategies as part of the game play.
- **CS_{9,2}**: Learning of new Machiavellian strategies.
- **BP_{9,1-2}**: The bot is able to reason about opponents' Theory of Mind, i.e. "I know you know I know" (Lewis, 2003). Therefore, it shows social intelligent behaviors like preparing an ambushade.



➔ Current Bot Implementations

Agent	L_2	L_3	L_4	CLS	CQS
<i>Reactive-Bot</i>	1	0	0.000	1.000	0.18
<i>Adaptive-Bot</i>	1	1	0.000	1.250	2.22
<i>Attentional-Bot</i>	1	1	0.216	1.255	2.38



➔ Conclusions

- Inferring BP from observations is sometimes misleading.
 - Use of testing protocols like BotPrize.
- Higher level behavioral profiles are difficult to test.
 - Levels 9 and 10 required extremely complex environments (real world).
- Same agent, different developmental stages over time.

➔ Thank you

