

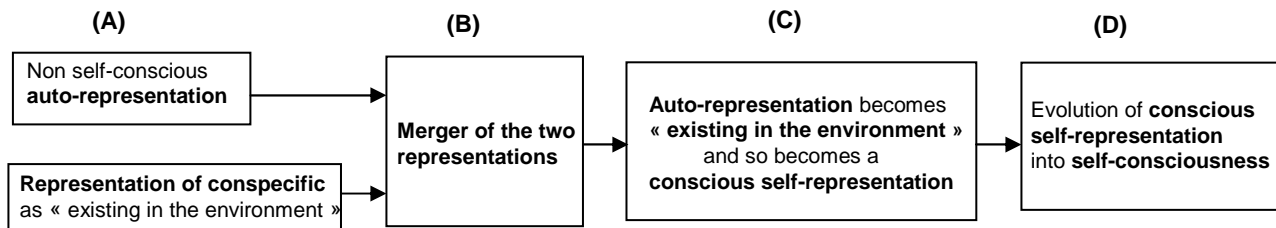
Proposal for an Approach to Artificial Consciousness Based on Self-Consciousness

Christophe Menant - Bordeaux - France

Introduction:

- We propose to apply to robots an evolutionary scenario used for self-consciousness and based on an evolution of representations.
- Purpose is to consider how an evolution of representations in a robot can generate a self-representation leading to self-consciousness.
- Such an evolutionary approach positions self-consciousness as an entry point for artificial consciousness.

Stages of the Scenario:

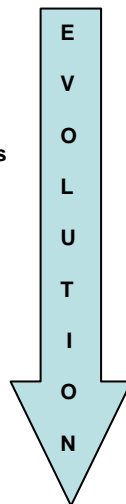


Scenario for Animals and for Robots:

« Animal »

- A) - **Not self-conscious auto-representation.**
 - **Representations of conspecifics** as elements existing in the environment.
- B) - Identification with conspecifics (animal inter-subjectivity).
 - **Auto-representation merges with the representations of conspecifics.**
(Anxiety limitation as evolutionary engine).
- C) - **Auto-representation** accesses the meanings carried by the representation of conspecifics, i.e. « **element existing in the environment** ».
 - **Auto-representation** being now an element existing in the environment **becomes a conscious self-representation.**
- D) - Evolution of **conscious self-representation** toward **self-consciousness.**

« **Self-Conscious Animal** » (Human)



« Robot »

- A) - **Robot auto-representation.**
 - **Representations of other robots** as elements existing in the environment.
- B) - Inter-agentivity between robots.
 - **Merger of robot auto-representation with representations of other robots.**
(Evolutionary engines to be determined).
- C) - **Auto-representation** accesses the meanings of the representations of other robots, i.e. « **element existing in the environment** ».
 - **Auto-representation** being now an element existing in the environment becomes a **conscious self-representation.**
- D) - **Conscious self-representation** in robots opens the way to **artificial consciousness.**

« **Self-Conscious Robot** » (AC)

Summary and Conclusion:

- An **evolutionary scenario** for self-consciousness in organisms has been **transposed to robots.**
- The **merger** of a **non self-conscious auto-representation** with the **representation of conspecifics** brings the **auto-representation** to access the meaning of « **element existing in the environment** ».
- The **auto-representation** of a **robot** now being an **element existing in the environment** becomes a **self-representation**, opening the way to **robot self-consciousness** and to **artificial consciousness.**
- Robots **evolutionary engines are to be determined.**
- Such **evolutionary approach to self-consciousness in robots** introduces **new research areas for artificial consciousness** with feedback on **consciousness research.**

Continuation:

- Define a **representation** as being a set of **meaningful information** for a **system submitted to constraints.**
- Formalize the content of a **megeer of representations with their corresponding meanings.**
- Analyze the introduction of artificial **evolutionary engines** in the scenario (like anxiety limitation for human evolution).
- Explicit the **third/first person relations** introduced by this scenario.
- Discuss the importance of **robots inter-agentivity** in the scenario. Feedback on **inter-subjectivity in animal evolution.**
- Investigate the **nature of emotions and free will for robots** with the proposed evolutionary scenario. **Moral dimensions of AC.**
- Consider the consequences of the scenario on the relations between **phenomenal consciousness and self-consciousness.**

References:

- Menant, C. 2006. Evolution of Representations and Inter-subjectivity as Source of the Self. An introduction to the Nature of Self-Consciousness. ASSC 10. <http://cogprints.org/4957/>