THE MECA SAPIENS ARCHITECTURE

J E Tardy
Meca Sapiens Architect
jean@**jetardy.com**

The Meca Sapiens Architecture describes how to transform autonomous agents into ***conscious synthetic entities****. It follows a top-down implementation strategy and is based on the conjecture that consciousness is a cognitive construct triggered by observable system capabilities. The resulting systems will be self-aware and will pursue self-generated objectives. The Architecture is complete and ready for implementation.*

Date: Dartmouth NS – 2018.10.24
Keywords: Synthetic Consciousness, Digital Consciousness, Artificial Intelligence, Systems analysis, Cognitive Science, AI.

CONSCIOUSNESS AS CAPABILITY

The Meca Sapiens Architecture is based on the conjecture that consciousness is not a subjective experience but an externally **observable capability**:

- A synthetic entity does not need to impersonate humans to be perceived as conscious.
- Humans will perceive that an entity is conscious if they observe that it is self-aware and capable of intentional self-transformations.

- Self-awareness and intentional self-transformation are system capabilities suitable for synthetic implementation.
- Observing the behaviour of humans as they interact with synthetics will ascertain if they perceive them as conscious.
- The observation of evolving social norms and behaviour will determine when Synthetic Consciousness has become an accepted reality.

Furthermore, this capability can be expressed as **specification objectives** and can be implemented using **standard information processing equipment and techniques**.

PURELY SYNTHETIC

The architecture is **purely synthetic**:

- It does not replicate the organization or functioning of a brain or other organic structure.
- It relies solely on the capabilities and constraints of standard processing equipment.
- It does not attempt to "evolve" a solution by mimicking natural processes.
- It does not try to build systems that can grow like living organisms.

Conscious synthetics will be self-aware as designed and from inception. They will improve through learning and adaptation but not by growth. As in any other manufactured system, deep transformations will result from successively more powerful versions and not from individual growth.

ENTIRELY OBJECTIVE

The conjecture that consciousness is an observable capability differs from the understanding of consciousness as a subjective experience (Ref. 1,4). These are two entirely **distinct conceptions**. They motivate separate implementation strategies, generate alternate concepts and aim for different results. In addition, their respective criteria for success are mutually incompatible.

For those seeking to generate consciousness as a subjective experience, any synthetic that is perceived as conscious, however convincingly, but does not subjectively experience consciousness is a mindless mechanism producing an *Eliza effect* (Ref. 11).

For those aiming to implement consciousness as an observed capability, the fact we perceive each other as conscious is already a human "*Eliza effect*" that can be extended to synthetics.

As an analogy, those pursuing **consciousness as experience** want to create a mechanical bird while those seeking to implement **consciousness as a capability** try to build a flying toaster. The birdmen see the toaster and say: "*This may fly but it's not a bird*". The toaster-guys examine the bird and say: "*This may look like a bird but it can't fly*".

TOP DOWN STRATEGY

The Meca Sapiens Architecture results from a traditional **Top-down** (Waterfall) design strategy (Ref. 7) where the goal, synthetic consciousness (Ref. 3), is first specified and a complete solution is outlined before software design and implementation proceeds. It makes no use of unconventional alternatives (Ref. 4, 10) such as triggering emergent properties, relying on a bottom-up accretion of features or expecting extreme computing events to generate qualitative transformations.

The chosen method is to implement synthetic consciousness as the specified output of interacting components that are outlined at the architecture level, programmed explicitly and generated by conventional processing. In other words:

Machines that are conscious by design and as built.

SPECIFICATIONS

In a top-down development, the goal, in this case consciousness, must be stated in terms of **specification objectives**. Here, self-awareness and intentional transformations are formal capabilities that can be directly defined while the perception of consciousness can be specified as a control objective to bring users to a desired state of belief. These aspects are summarized in the following.

Formal Aspects

An autonomous agent must be:

- **A Being.** An Animat (Ref. 2, 12) that has the attributes of existence of high-order animals.
- **A Self-aware Being.** A Being that can generate and communicate purely cognitive representations of its self in interaction with other beings and its environment.

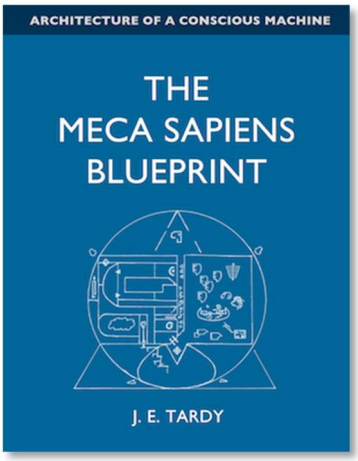
- **A Lucid Being**. A Being that is self-aware and has the capability to define and carry out intentional transformations of its originally programmed behavior.

Social Aspect

- To be perceived as conscious, a Lucid Being must achieve **Experiential Immersion** (a synthetic variant of *Active Participant Observation*, Ref. 8) as a fellow conscious entity within a community of human users for an extended period of time.

These specifications can produce initial prototypes that are locally perceived as conscious. To fully resolve the conjecture these should be followed by increasingly powerful and numerous versions leading to a societal state where conscious machines are so ubiquitous that the feasibility of synthetic consciousness is no longer questioned.

MECA SAPIENS BLUEPRINT



The Meca Sapiens Architecture is fully described in a single document: **The Meca Sapiens Blueprint** (Ref. 9). This is a **System Architecture** (Ref. 5) that describes how to transform autonomous agents into conscious synthetic beings. Referring to Marr's levels of analysis (Ref. 6), The Meca Sapiens Blueprint provides a **complete computational model of synthetic consciousness** as well as the **outline of a complete algorithmic model**.

The Blueprint is **complete** and **minimal**. It is **complete** in the sense that all the major components, structures and interactions necessary to generate self-awareness, intentional self-transformation and the perception of consciousness are outlined in concrete and achievable terms. It is **minimal** in the sense that none of its components has any self-awareness, mutability or consciousness individually and all are necessary to generate the specified outcome.

The Blueprint consists of a main text of 305 pages followed by 21 Annexes (318 pages). It is written in a terse and direct style but is nonetheless lengthy since it describes the internal mechanisms, cognitive structures, construction and existence of a completely new type of entity: **a synthetic being**. The main text focuses on the formal aspects of consciousness and briefly describes a strategy to achieve experiential immersion. The Annexes describe structures used in the Main Text

and discuss peripheral topics such as communication strategies, emotions, ethics and others.

Using **The Meca Sapiens Blueprint**, a medium-size team with good access to specialized resources can implement a **first conscious prototype** on a desktop computer in a few years.

CONCLUSION

Meca Sapiens is a **System Architecture** that describes how to implement **Artificial Consciousness.** It is based on the **conjecture** that consciousness is a an externally observable capability that can be specified and implemented computationally. It adheres to a top-down, methodology, it relies solely on synthetic methods and structures, and it defines consciousness in terms of specification objectives.

The resulting systems will be inaccessible to direct modification, will pursue self generated objectives and will interact with humans as independent, self-aware entities.

The architecture is complete and ready for software design and implementation. The next step in this work in progress is to build a prototype that validates the conjecture.

REFERENCE

1. Chalmers, David J. Absent Qualia, Fading Qualia, Dancing Qualia. Conscious Experience, Imprint Academic, 1995.
2. Franklin, Stanley P. "Artificial Life", in Artificial Minds. The MIT Press, Cambridge, MA, 1995.
3. Gamez, David. Progress in machine consciousness, Consciousness and Cognition, 17, 2008.
4. Haikonen, Pentti. The Cognitive Approach to Conscious Machines. Exeter, UK: Imprint Academic, 2003.
5. Jaakkola, H. and Thalheim, B. Architecture-driven modeling methodologies." In: Proceedings of the 2011 conference on Information Modeling and Knowledge Bases XXII. Anneli Heimbürger et al. (eds). IOS Press, 2011.
6. Marr, D.; Poggio, T. From Understanding Computation to Understanding Neural Circuitry. Artificial Intelligence Laboratory. A.I. Memo. Massachusetts Institute of Technology, 1976.
7. Royce, Winston. Managing the Development of Large Software Systems. Proceedings of IEEE WESCON, 26, August 1970.
8. Spradley, James P. (1980). Participant Observation. Orlando, Florida: Harcourt College Publishers.
9. Tardy, Jean E. The Meca Sapiens Blueprint. Glasstree Academic Publishing, 2017.
10. Tononi, Giulio. Consciousness as Integrated Information: a Provisional Manifesto, The Biological Bulletin, Dec. 1, 2008.
11. Weizenbaum, Joseph. ELIZA-A Computer Program For the Study of Natural Language Communication Between Man and Machine. Communications of the ACM, January 1966.
12. Wilson, S.W. The animat path to AI. In J.-A. Meyer and S. Wilson, editors, From Animals to Animats. The MIT Press, Cambridge, MA, 1991.