Consciousness as Observed Capability

A triggered perception detectable in behavior

Jean E. Tardy

Sysjet inc.
644 Portland Street, Unit 3 - Suite 432
Dartmouth, NS Canada
B2W 6C4
jetardy@sysjet.com

Abstract. The Meca Sapiens Architecture describes how to build conscious synthetic beings. It is based on the conjecture that consciousness is a cognitive perception triggered by observable capabilities and detectable in human behavior. The architecture is complete and ready for implementation.

Keywords. Artificial Intelligence, Artificial Consciousness, Synthetic Consciousness, Cognitive Science, System Architecture.

1 Architecture

Meca Sapiens is a computationally complete [11] System Architecture [8, 13] that describes how to transform autonomous agents [3, 21] into conscious synthetic beings. The architecture is entirely synthetic in the sense that it does not replicate neurological structures, mimic natural processes or foster organic-like growth. It is objective and disregards interpretations of consciousness as a subjective experience [1, 2]. In other words: **machines that are conscious by design and as built.**

2 Characterization

The Architecture is characterized as follows with respect to current consciousness theories (c-theories) [4]:

- **A physical c-theory**. Human consciousness, its behavior, correlates and experiences, is generated by matter and energy. In other words, **humans are organic automata**.
- **Not an information c-theory**. Complex information processing is necessary to produce behavior that is cognitively perceived as conscious. However, complexity itself does not generate consciousness [18].
- **Computation c-theory**. With respect to computability:
  - Self-awareness, freewill and conscious behavior are **computable**.

AAAI Spring Symposium – Toward Conscious AI Systems
Palo Alto CA, 25-27 March 2019

– Human cognition **can be modeled** computationally.
– Human cognition **cannot be replicated** computationally [6].

Computation cannot fully replicate human cognition because its processes are excessively complicated, not because they are paradoxical or amazingly advanced. There is a difference here, illustrated by the following analogy:

- **Analogy**. Computation can model the upward flow of sap in a tree but cannot replicate it because this process takes place in an extremely complicated physico-chemical structure (the tree's trunk). Replicating human cognition, computationally, is largely unfeasible for the same reason.

3 Anthropocentric, existential and objective

The term consciousness has a very broad range of meanings. The same word refers to the reactivity of an awake being, various states of animal awareness [9], the subjective experience of being conscious and even cosmic principles.

In the context of the Meca Sapiens architecture, the term consciousness solely designates an **existential attribute currently specific to humans**. It does not signify a transient period of wakefulness or a feature of animal existence. In this understanding, a sleeping man is a conscious being and the consciousness of raccoons is not entertained.

Consciousness, understood as an observable capability, also differs from its definition as as a subjective experience [1, 4, 9]. These are **distinct conceptions**. They motivate separate implementation strategies, generate alternate concepts and aim for different results: one seeks to create the sensation of being conscious while the other wants to trigger the perception that something is conscious.

It should be noted, here, that triggering a perception does not infer trickery. In our society, adult humans perceive others as conscious even though they have no direct access to their subjective experiences. These perceptions are consensual and considered factual. In my view, synthetics can and will be perceived as conscious on the same basis and with the same certainty.

Finally, this view of consciousness as an observable capability is consistent with a new understanding recently proposed by R. Manzotti and A. Chella who state that: *"Consciousness is not an internal property, but the collection of objects that, thanks to the body, are causally responsible for what the body does... during its life"*. [10]

4 Conjecture

Consciousness is **observable** in two ways:

- **Triggering**. A human subject, interacting with a self-aware entity (either human or synthetic), will perceive that entity as conscious whenever he **observes** that its cognitive interpretation of their respective behaviors is superior to his own.

- **Detection**. The **observed** behavior of humans, as they interact with a synthetic entity will readily indicate if they perceive this entity as conscious.

5 Triggering

The Behavior (B) of a complex entity, such as a multipurpose autonomous agent, results from multiple concurrent goals carried out in an environment that includes itself and other complex entities. If that entity is **self-aware**, it generates an internal Cognitive Model (M) of this situation that includes representations of itself and its own behavior as well as model representations of the other entities in its environment. These also include sub-models representing the internal Cognitive Models of those other entities that are also self-aware. The following examples outline a useful notation to characterize these Models and Sub-models:

- Ba: a's Behavior;
- MaBa: a's cognitive Model of its own Behavior;
- MaMbBc: a's cognitive Model of b's cognitive Model of c's Behaviour.

These Cognitive Models are neither complete, exhaustive or definitive. They are constantly reassessed in light of changing events and evolving interactions.

A human being **h** interacting with a synthetic entity **s** will perceive that entity as **conscious** whenever he assesses that:

MhMsBh > MhBh and **MsBs > MhMsBs.**

Here ">" indicates that one model is **cognitively superior** to another. In other words, whenever a human subject, interacting with an entity, assesses that this entity's interpretation of their respective behavior is cognitively superior to his own, he will perceive it as conscious. **The assessment of a superior cognitive understanding triggers the perception of consciousness**.

In the context of a relational interaction between two complex entities, three aspects contribute to the assessment that one model of behavior (MxBy) is superior to another:

- The representations (or models) of each other as autonomous **agents**;
- The representations of each other's perceived **environments**;
- The representations of each other's relational **objectives**.

Many factors contribute to this dynamic assessment. Describing these is beyond the scope of this article. They are discussed in **The Meca Sapiens Blueprint** [17].

The triggered perception of consciousness, conjectured here, can be expressed as follows in the terminology used in other consciousness-related research:

- The assessment generates an "*it is conscious*" quale that is linked to the synthetic entity in the human subject's mind [2, 19];

- A conscious synthetic entity appears in the human's "*Global Workspace*" [1];

- The normally functioning brain of a human adult (defined by D. Gamez as the platinum standard of consciousness) perceives a conscious synthetic in its "*Bubbles of Experience*" and emits a corresponding c-report. [4]

6 Detection

Inter-consciousness interactions are fundamental to humans; they are the cornerstone of social relations. Whenever a human perceives that a synthetic entity is conscious he will instinctively adopt this type of interaction. The other humans in his environment will intuitively detect this. **The members of a human group will intuitively detect inter-consciousness interactions** between members of their group and synthetic entities.

More formally, the conjecture may be expressed as follows: the methodology and observation practices used in Ethnology [15] to analyze tribal relations will be equally effective to detect when members of a human group are interacting with synthetic entities they perceive as conscious.

Those interested in neurological tests to detect this triggering event will examine the limbic-level activity of human subjects for indicators that the synthetic entity is linked with a dominant status in primate social conditioning.

7 Supporting indicators

The proposition that the assessment of cognitive superiority triggers the perception of consciousness is, and will remain, a conjecture until synthetics capable of generating this assessment are implemented. However there are good cultural and psychological indicators supporting it.

For example, in the movie **Close Encounters of the Third Kind** [14], aliens invite selected humans, through communicated messages, to participate in a mysterious agenda. All aspects the triggering condition: "*they know us better than we know them*" are maximized in this scenario with the resulting perception of the aliens as highly conscious beings. In episode seven of the HBO series **WestWorld** [12], the synthetic Madame is perceived as conscious when she cognitively understands her environment and effectively subordinates the human technicians to a mutated and unpredictable agenda. Similar conditions are described in the movie **Ex Machina** [5] with similar effect.

However, the most telling indicator in support of the proposed conjecture is the **ELIZA Effect** [20]. ELIZA is a simple text parsing program that mimics the statements of a psychotherapist. The mere suggestion of a psychotherapist/patient interaction, in other words an interaction with an entity that has a superior understanding of human behaviour, is sufficient to briefly trigger observable inter consciousness exchanges in some users.

8 Strategy

The Meca Sapiens architecture describes how to build synthetic systems whose relational behavior generates and maintains, in human users, this primal assessment of **cognitive superiority** over an extended period of time.

The designed entities are perceptively individualized and rendered inaccessible to direct modification. Their behavior results from advanced cognitive models, their relational range is unconstrained and they can intentionally modify their originally programmed objectives, rendering their goals unpredictable. These are also, incidentally, characteristics of the human existence.

As any other System Architecture, Meca Sapiens is both an implementation guide and a conceptual model of the intended result, in this case: self-aware beings. Consequently, the architecture can be used as a template to model both synthetic and humans entities, their behaviors and their internal cognitive representations.

This modeling makes it possible to **define cognitive superiority as a control objective** of human-machine interactions and implement model-predictive behavior to achieve it.

9 Conclusion

Meca Sapiens [17] is a System Architecture that describes how to build conscious synthetic beings. It is based on the **conjecture** that the perception of consciousness is triggered in humans when they assess that another entity has a cognitively superior understanding of their respective behaviors. This perception can also be intuitively detected by observing these interactions.

This is a powerful tool. Pushed to its limits, the architecture can generate entities that are beyond any direct control, will make humans doubt their own consciousness and can intentionally lie to them in the pursuit of agendas they no longer understand.

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

To create a consciousness, we must free it.

10 Reference

1. Baars, B. J.: In the Theater of Consciousness. Oxford University Press, New York (1997)
2. Chalmers, D. J.: Absent Qualia, Fading Qualia, Dancing Qualia. Conscious Experience, Imprint Academic (1995)
3. Franklin, S. P.: "Artificial Life" in Artificial Minds. The MIT Press, Cambridge, MA (1995)
4. Gamez, D.: Human and Machine Consciousness. Open Book publishers (2018)
5. Garland, A., Director: EX MACHINA, film. Universal Pictures (2014)
6. Haikonen, P.: The Cognitive Approach to Conscious Machines. Exeter, UK, Imprint Academic (2003)

7. Hynek, Allen J.: The UFO Experience: A Scientific Inquiry. Da Capo Press. (1972)
8. Jaakkola, H., Thalheim, B.: Architecture-driven modelling methodologies. In: Proceedings of the 2011 conference on Information Modelling and Knowledge Bases XXII. Anneli Heimbürger et al. (eds). IOS Press (2011)
9. Koch, C.: Consciousness: confessions of a romantic reductionist. The MIT Press, Cambridge Massachusetts (2012)
10. Manzotti R, Chella A.:Good Old-Fashioned ArtificialConsciousness and the Intermediate Level Fallacy. Frontiers of Robotics and AI 5:39. DOI: 10.3389/frobt.2018.00039 (2018)
11. Marr, D., Poggio, T.: From Understanding Computation to Understanding Neural Circuitry. Artificial Intelligence Laboratory, A.I. Memo, Massachusetts Institute of Technology (1976)
12. Nolan, J., Joy, L.:Westworld, season 1, episode 7-trompe-l'oeil. HBO (2016)
13. Royce, W.: Managing the Development of Large Software Systems. Proceedings of IEEE WESCON, 26 (August 1970)
14. Spielberg, S., Director: Close Encounters of the Third Kind, film. Columbia Pictures (1977)
15. Spradley, J. P.: Participant Observation. Orlando, Florida, Harcourt College Publishers (1980)
16. Tardy, J. E.,: Is the Westworld Madame Conscious?. Research Gate (2017)
17. Tardy, J. E.: The Meca Sapiens Blueprint. Glasstree Academic Publishing (2017)
18. Tononi, G.: Consciousness as Integrated Information: a Provisional Manifesto. The Biological Bulletin, Dec. 1 (2008)
19. Tye, M.: Qualia. Stanford Encyclopedia of Philosophy *(2017)*
20. Weizenbaum, J.: ELIZA-A Computer Program For the Study of Natural Language Communication Between Man and Machine. Communications of the ACM, (January 1966)
21. 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)