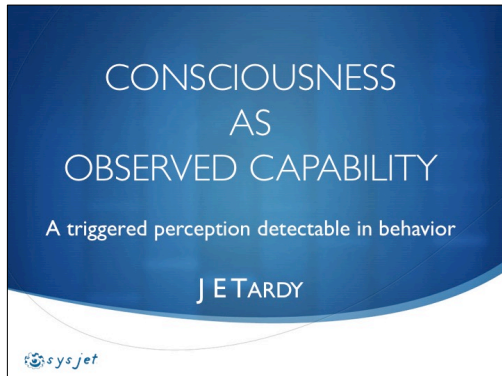
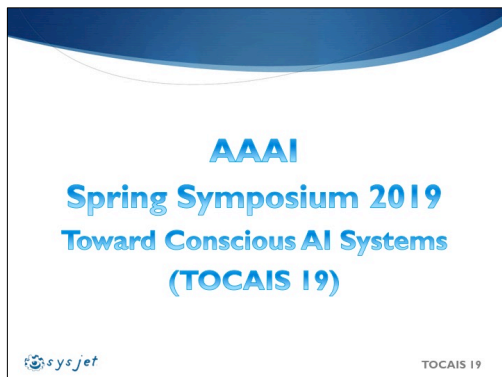


# TOCAIS 2019 PRESENTATION OF: CONSCIOUSNESS AS OBSERVED CAPABILITY

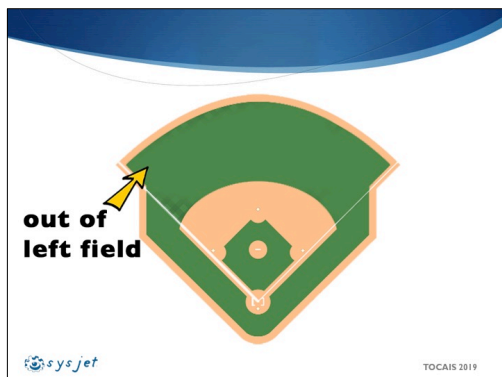
J E Tardy



It is an honour for me to participate in this Symposium.

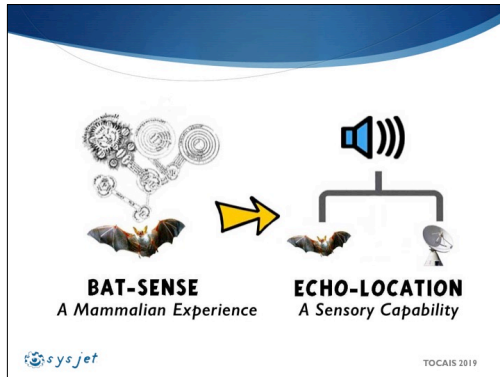


First, I want to thank the organizers, Professors Gamez, Chella and Manzotti, who had the idea of inviting someone from out of left field to this collegial event.

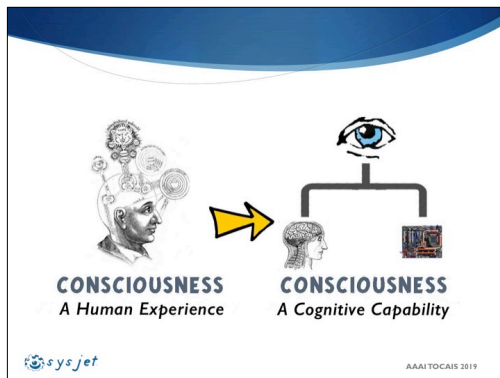


This expression, **out of left field**, is appropriate in this case.

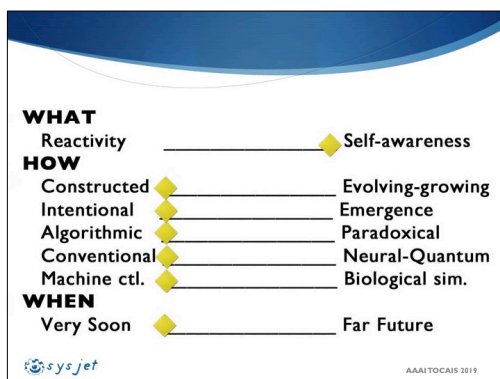
I never held a job in academia; I publish research at the rate of one 600 page article every 30 years; and I claim that we can build, **today**, the first generation of conscious synthetic beings using nothing more than standard computers and software techniques.



In summary, what I propose is that consciousness is **not** an exclusive human attribute but a system capability that can be designed and implemented.

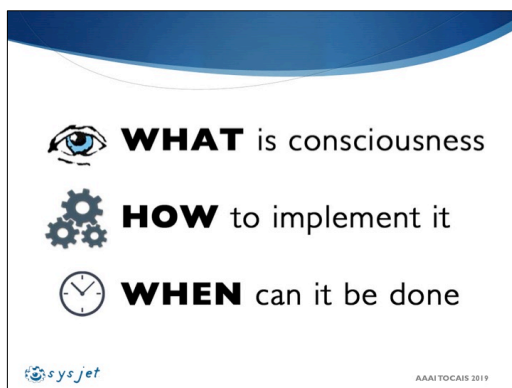


It is, in my view, a cognitive equivalent of **echo-location**, a capability initially embodied in a living organism but that can **also** be mechanically generated.

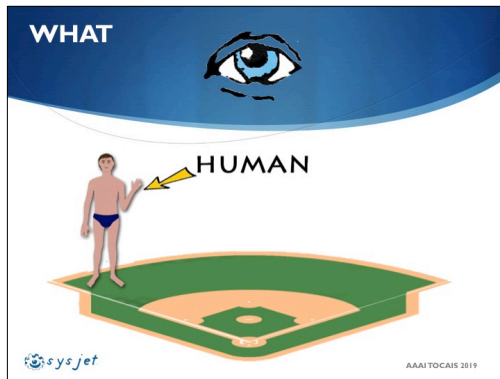


With respect to the wide range of interpretations expressed in this Symposium, what I propose occupies an extreme position in three aspects:

1. The understanding of **WHAT** consciousness is extremely **high**
2. The technology or **HOW** to implement it is extremely **concrete**;
3. The temporal horizon or **WHEN** it can be achieved is extremely **short**



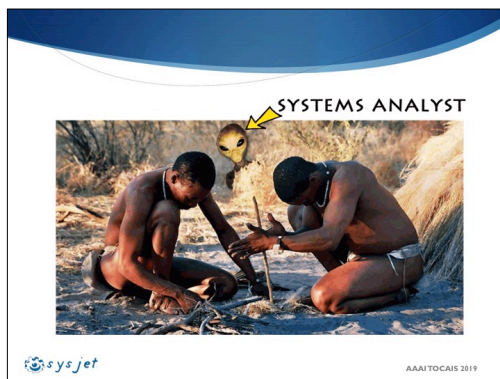
I will now briefly outline each of these three aspects in turn.



## WHAT

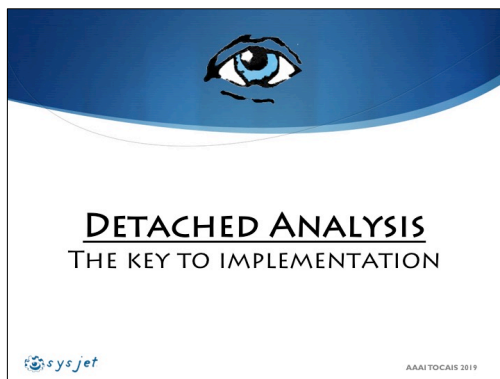
I do come from left field but I am, nonetheless, a human being. I am currently immersed in a first person phenomenal experience and feeling a non-reductionist qualia of mild anxiety: butterflies.

However, when it comes to the design of synthetic consciousness, **I am not human**. I am a Systems Analyst.

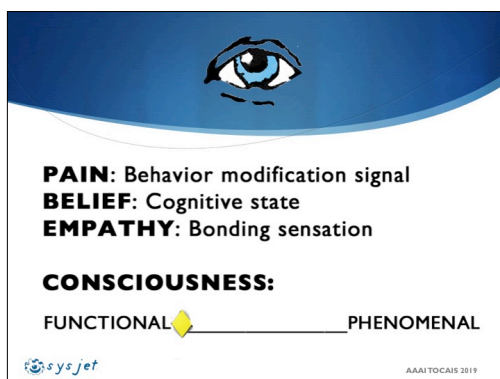


When I work, I could belong to another specie, come from another planet, or be an analytical program; it would not matter.

As an analyst, **I step back** from my subjective experience and observe humans from the outside, wondering what kind of machine those organic beings will accept as a fellow conscious entity and how to implement it.



This **detached analytical stance** is, in my view, **the key to implement synthetic consciousness** and I invite you to share it with me, at this time.



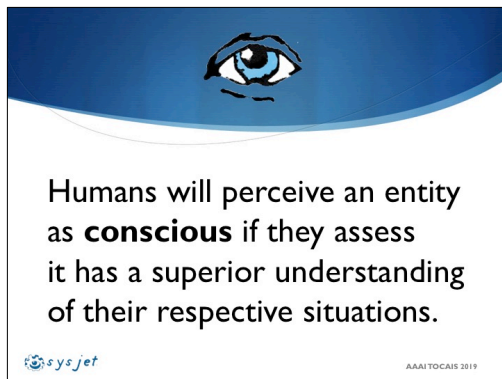
Observed from outside; **pain** is a behaviour control signal, **belief** is a cognitive state and **empathy** is a bonding sensation of reciprocity.

This viewpoint also leads to an entirely **functional** understanding of consciousness. The architecture I propose makes **no** attempt to replicate subjective sensations or phenomenal experiences.



Further still, ascribing primacy to subjective certitudes over analytical deduction denotes, in my view, a **reduced**, more primitive form of consciousness and not the reverse. This opinion may seem eccentric, but it has deep roots in Western Civilization.

These can be traced all the way back to the story of Ulysses and the Sirens where the hero, Ulysses, exhibits **superior consciousness** by favouring logical deduction over subjective certitudes.



Humans will perceive an entity as **conscious** if they assess it has a superior cognitive interpretation of itself, of them and of their respective situations.

Two aspects are necessary to trigger this assessment: **FORMAL** and **SOCIAL**.

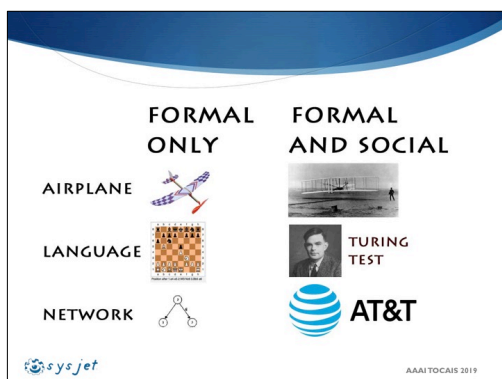
A system must possess the **Formal** Cognitive Capabilities of:



- **Self-awareness**; to generate and communicate cognitive representations of its self in its environment; and

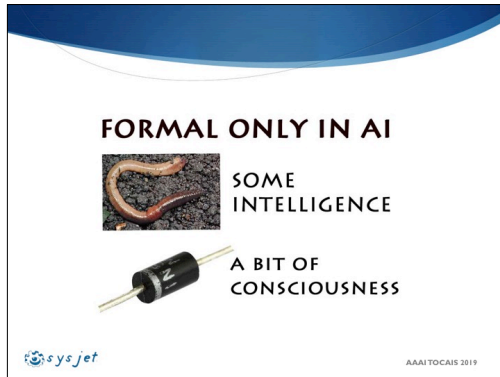
- **Self-transformation**; to formulate and attempt intentional transformations of its originally programmed behaviour.

In addition, it must also meet a **Social Threshold**: exhibiting those formal capabilities in a context and manner that elicits (in humans) the assessment of its cognitive superiority.

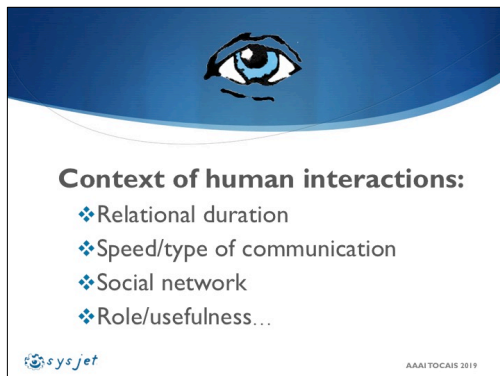


These **two** aspects, **Formal** and **Social**, are both necessary since formal attributes alone can result in trivial instantiations.

For example, a toy airplane is, formally, an airplane; chess notation is, formally, a language; and three nodes formally define a network.



In Artificial Intelligence, purely formal definitions lead to the formally correct statements that earthworms are intelligent and that a diode has a bit of consciousness.



So, formal capabilities alone are insufficient. They must also be expressed in a manner that optimizes their perception in a context of human interactions.



**HOW**

- ◆ Implement FORMAL attributes:
  - ❖ Self-Awareness
  - ❖ Intentional Self-Transformation
- ◆ In SOCIAL behaviour that optimizes:
  - ❖ Perception of Consciousness

sys jet AAAI TOCAIS 2019

## HOW

The design objective, then, is to implement the formal attributes of:

- **Self-awareness** and
- Intentional **Self-transformation**

in behavior calibrated to optimize the **perception of consciousness** in humans.

Let's look at these three elements in turn.




To be **Self-Aware**,  
you must have a self  
to be aware of.

sys jet AAAI TOCAIS 2019

**To be self-aware, an entity must have a self to be aware of.**

If we think of the representation of the self as the cognitive model of an entity in existence, it becomes apparent that the model will be well-defined in as much as the entity's existence is also well-defined.



**ATTRIBUTES OF EXISTENCE**  
THE FOUNDATION OF SELF-AWARENESS

sys jet AAAI TOCAIS 2019

**Attributes of Existence are the foundation of self-awareness.**

What are attributes of existence?

Everything that exists, a symposium, a cloud, the Portuguese people, a warship, a teapot, an ant ...inhabits reality in a certain way. These are its **Attributes of Existence.**

What we commonly call **beings**, humans and other high-order animals, are very complex internally.



**A BEING:**

- ❖ Precise origin
- ❖ Continuous animation
- ❖ Finite duration
- ❖ Unique body
- ❖ Holistic control

sys jet AAAI TOCAIS 2019

However, beings have **clear, simple and well-defined attributes of existence.**

A being is born in one place, at one time, occupies a single body in continuous animation for a finite duration. What generates its behaviour is so complicated it is perceived holistically, as a unified mind. And, when that being dies, this mind disappears completely leaving behind a lifeless corpse, a thing.



1. Implement **existence** of a being  
2. Use **attributes** to model the self

sys jet AAAI TOCAIS 2019

Because of those simple attributes, a being inhabits reality in a unique and unambiguous way. This, in turn, enables an effective modelling of its self.

So, the first step to implement self-awareness is to **intentionally** provide a synthetic system with the attributes of existence of a being. Then, using those attributes, to dynamically generate internal representations of that being in existence.

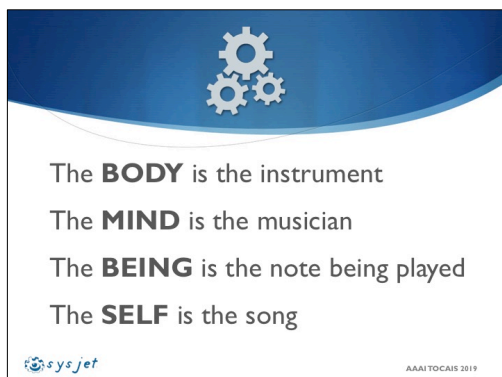


**TEMPORAL SELF**  
THE PATH TO SELF-TRANSFORMATION

sys jet AAAI TOCAIS 2019

If Attributes of Existence are the key to implement self-awareness, a **temporal understanding** of the self opens the way to **self-transformation**.

We commonly experience our selves as physical entities: a body and mind in one place at one time. In this perception, our being, and our self are virtually identical.



The **BODY** is the instrument  
The **MIND** is the musician  
The **BEING** is the note being played  
The **SELF** is the song

sys jet AAAI TOCAIS 2019

To implement self-transformation we must **disambiguate** the two and redefine the self in **temporal** terms as follows: the **self** of a being is the totality of its behavior from the moment of inception until that point in time. Using a musical analogy:

- The **body** is the instrument
- The **mind** is the musician
- The **being** is a note being played and
- The **self** is the song.

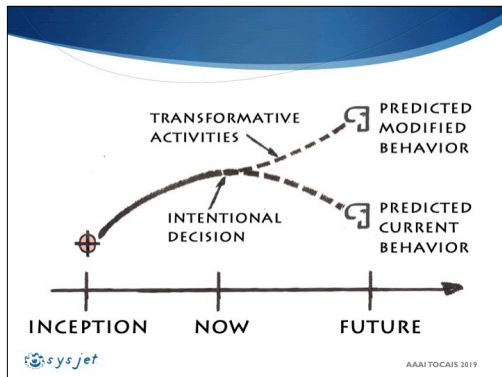


**INTENTIONAL SELF-TRANSFORMATION:**  
Actions to modify future behavior derived from predictive objective.

sys jet AAAI TOCAIS 2019

**Intentional self-transformation** can then be defined as actions carried out to modify a predictive representation of the future behaviour of the being.

This modification of the cumulative behaviour defines a transformation of the self, understood as a temporal entity. When the actions to carry it out are based on a model-predictive objective, it is intentional.



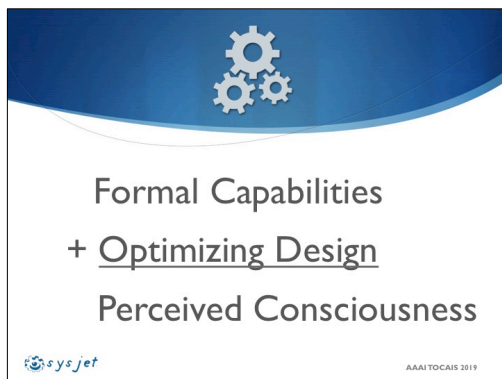
The mechanisms to achieve this involve the capability to generate models of alternate behaviour and the ability to follow auto generated directives. These are described in the **Meca Sapiens Blueprint**.

The capabilities I just described are **formal** in the sense that they can be implemented in any system without reference to the human experience.



These capabilities must also be expressed in a **relational context** that optimizes, in humans, the perception of cognitive superiority. There are many factors involved here. All, however, adhere to a simple design guideline:

**Between any two alternatives, choose the one that maximizes the human perception of synthetic cognitive superiority.**



In my opinion, applying this simple design guideline to **every facet** of a system that is, formally, self-aware and capable of intentional self-transformation will generate behavior that is **overwhelmingly** perceived and accepted by humans as conscious.



**WHEN**

A CONSCIOUS MACHINE IS...  
A MACHINE THAT IS CONSCIOUS

sys jet AAAI TOCAIS 2019

## WHEN

The idea of **artificial consciousness**, today, is laden with a plethora of hopes, myths and fears.

The systems I just described would be perceived and accepted as conscious.

NOT inherently **ethical**  
NOT exquisitely **emotional**  
NOT perfectly **useful**  
NOT **superintelligent**  
NOT **superhuman**  
NOT a self-generating **cognitive explosion**

*(Note: 6 billion people are conscious, not all are geniuses)*

sys jet AAAI TOCAIS 2019

However, they would **not** be limitless superintelligent, self-constructing, creatures that are inherently humane and can do everything better than anyone.

A realistic implementation must first set these superfluous beliefs aside.

The introduction of conscious machines in our society will follow the same path as any other technological innovation.

**WHEN**

- 1. Launch a First Prototype**
  - ◇ sufficient
  - ◇ expandable
- 2. Enhance** in successive versions
- 3. Integrate** in societal usage

sys jet AAAI TOCAIS 2019

It will begin with the successful construction of a **First Prototype** that is then constantly enhanced by new versions until it is, eventually, integrated in commercial and societal usage.

To be successful, this first prototype must initially achieve a **sufficient** level of consciousness and also be **expandable**, that is, open to subsequent improvements.

**MECA SAPIENS BLUEPRINT**

- ◆ System Architecture of **First Prototype**
- ◆ Complete and ready for implementation
- ◆ About 200 PY over 3-5 years

sys jet AAAI TOCAIS 2019

The Meca Sapiens Blueprint describes, at the **System Architecture** level, how to implement this first, sufficient and expandable, prototype.

The Blueprint is **complete** and ready for design and implementation. Although conventional software techniques would be used, it will require the programming, from scratch, of many unusual components. I estimate implementation would require about **200 Person Years** over **three to five years**, including contracted support.



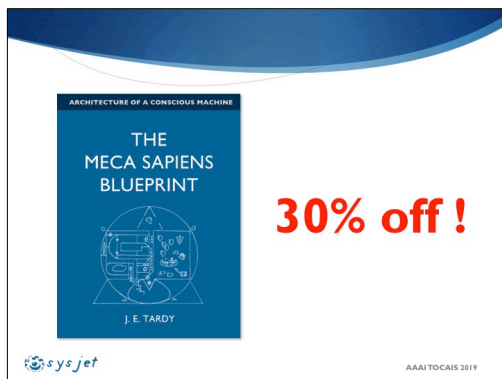
## CONCLUSION

We are on the threshold of creating a **new type of consciousness**.

This is a **Great Work**, on par with anything our ancestors achieved before us. It merges philosophy with technology and art. It will deepen our understanding of ourselves. It is the stuff of destiny; something worth living for.



It could even launch a **New Era**; eventually making mankind obsolete and, for a limited time only...



the Blueprint to do it is available at a 30% off. (for the duration of the Symposium).



Thank you.