CONSCIOUSNESS AS OBSERVABLE CAPABILITY

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

*Meca Sapiens defines consciousness as an **observable capability** suitable for standard implementation. Humans will recognize a system as conscious if they observe it has self-awareness and intentionality. If, as suggested, a conventional development is sufficient to generate artificial consciousness then the first prototypes will trigger an explosive proliferation of conscious machines.*

AIM AND PATH
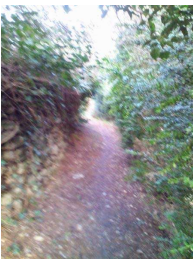
The destination decides the path.

This statement, understood in terms of system implementation, means that the end-result the builders foresee determines the development strategy they adopt. If they understand the objective analytically, as the outcome of well-defined interacting compo-

nents, they will pursue a systematic construction process to achieve it. If, on the other hand, they perceive the goal in terms of a holistic subjective experience, they will seek to create underlying conditions from which the desired result can arise.

The first will implement and connect components to produce an intended result. The second will accumulate diversified processes and interactions to trigger an emergent effect.

The first will attempt to build a consciousness directly. The second will assemble and activate a substrate from which, they hope, consciousness will appear.

Those who view digital consciousness as a construction effort will follow a proto-typing process, beginning with simplified subsystems and basic processors to implement initial prototypes then iteratively expanding these first versions with increasingly complex components. Those seeking to generate emergent properties will run ever more complicated interactions on increasingly powerful computers until the desired event appears.

In my view, under a technical veneer, the quest for emergent properties reflects a primitive mindset. This approach spawned notions such as the belief that complexity spawns consciousness or that an electric brain could have human sensations. Such ideas belong to a pre-technical era.

SUBJECTIVE PERCEPTIONS

 Our current understanding of consciousness is derived from traditional philosophical concepts that predate information processing. They were fashioned at a time when humans were the only entities capable of communicating an understanding of themselves and of the world. This situation blurred the distinction between subjective human experiences and events that are independent of the human condition.

In particular, the human experience was (and still remains) the only available source of information about consciousness. As a result, the prevalent understanding of consciousness remains centered on the subjective human experience.

The following two typical definitions of consciousness highlight this human specific bias. One definition states that:

> *"Consciousness is the fact of awareness by the mind of itself and the world."*

Where, the term "mind" is commonly defined as: a human being's reasoning and thoughts.

Another common definition of consciousness states that:

> *"Consciousness is the upper level of mental life of which the person is aware as contrasted with unconscious processes."*

Here again, the term "person" refers to a human person. Also, the "awareness" mentioned in both definitions is often described, in philosophical discourse, in terms of "qualia" which are themselves understood as individual instances of subjective conscious experience.

Given the heavily subjective character of these and other understandings of consciousness prevalent today, it is not surprising that so much effort has been invested in concocting emergent conditions rather than attempting an intentional construction. This resulted in thirty years of sterile pursuits.

BAT-SENSE

The first step toward a successful synthetic implementation of consciousness is to dissociate this concept from our subjective experience and redefine it in terms of system capabilities.

This process transforms a unique and mysterious property that is present in a single organism into a system capability, whose mechanisms are understood and can be present in various entities. Another property, **echo location**, provides a good analogy of this evolution.

A few centuries ago, humans observed that bats could navigate in complete darkness without any apparent sensory input to guide them. At first, bats were the only creatures known to have that capability. This "*bat-sense*" was perceived as unique and very mysterious. It was the subject of various and at times fantastical interpretations. Could the bats directly feel the ether? Were they under demonic control? Was it bat magic?…

Eventually, the mechanism that it allowed the bats to navigate in the dark was discovered. The bats would emit sounds and perceive obstacles by sensing the reflection of those sounds. What was an utterly mysterious power became a new sensory mechanism that could be rationally explained and understood. Later, the ability to use echo sensing was also attributed to whales and what had been a bat-specific attribute, became **echo-sensing**, a feature common to various living organisms.
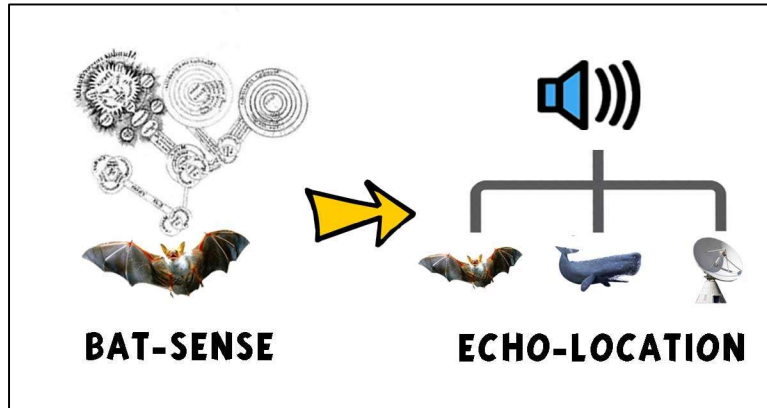
Finally, in the last century, humans developed synthetic systems, radars and so-nars, that were also capable of using reflected sound waves to perceiving their surroundings. The sensory feature possessed by a few living organisms became **echo location**, a system capability that could be detected in living organisms and could also be implemented in man-made systems.

CONSCIOUSNESS - A NEW BAT-SENSE

The prevalent understanding of consciousness, today, is similar to the notion of bat-sense a few centuries ago: a mysterious and unique attribute that only humans possess. As in the case of bat-sense, the mechanisms that generate the mysterious property of consciousness are the subject of fantastical conjectures: cosmic messaging, quantum effects, emergent properties, incomputable processes.

In addition, humans observed echo-sensing in a separate specie while consciousness is a human-specific attribute that is perceived subjectively. In the case of bat-sense, the mechanisms were mysterious but the capability was externally observable. With respect to consciousness, the causal mechanisms are mysterious **and** the capability is perceived subjectively. A double confusion that results in bizarre endeavors such as building synthetic brain-like structures to experience the subjective "qualia" sensations felt by humans.

CONSCIOUSNESS AS CAPABILITY

To implement artificial consciousness, our understanding of consciousness must undergo a transformation that is similar to what occurred with echo-location. An attribute that is currently perceived as a mysterious subjective experience exclusively shared by humans must become an **observable capability** that can be present in various systems and whose generating mechanisms are understood.

> An observable capability is a system capability that is independent of a specific system configuration and whose presence can be externally detected.

Moving from consciousness as experience to consciousness as observable capability means stepping away from the subjective sensations associated with consciousness and examining it as a human perception that is present under various conditions.
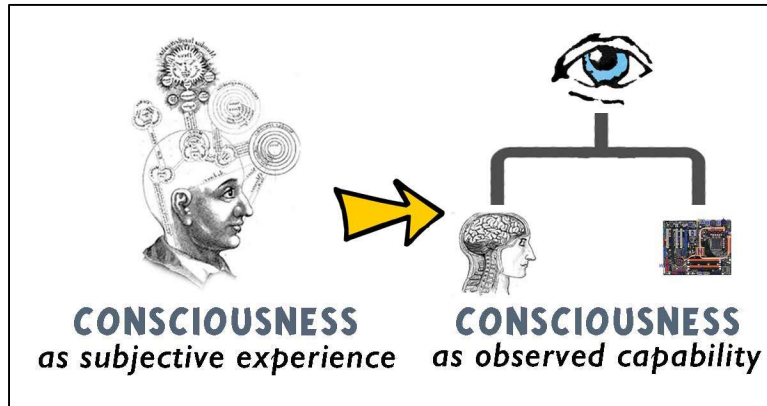
This shifts the **understanding** of consciousness away from subjectivity; going from from "*what is it to feel conscious*" to "*why does one person perceive another to be conscious*".

It moves the **objective** of artificial consciousness away from the subjective experience; going from "*how can we build a machine that feels conscious like we do*" to "*how can we build a machine that we, humans, will perceive as conscious*".

It transforms the **strategy** to implement consciousness from…

   Shared subjectivity: generating a synthetic emergence that shares the human subjective experience of consciousness.

To…

   Triggered subjectivity: building a control system that triggers, in humans, the subjective perception that an entity is conscious.

PERCEPTION VS REALITY

The point of view that consciousness is a subjective experience is self justifying and self-sufficient. Arguments cannot disprove it. Those who decree that "*consciousness is what we feel it is* " or *"consciousness is the way we feel"* will never be in error since their subjective experience is both the event and its measure.

A concept based on a subjective perception cannot be countered. A consensual opinion resulting from shared sensations cannot be challenged. The tenants of consciousness as experience can easily shrug off any proposal to implement consciousness as an **observable capability** by declaring that the outcome would be inconsequential:

> *"You may implement a system that appears to be conscious. It may even convince some people that it is conscious. However, such a construction will never be anything more than an animated puppet since it will not experience our inner subjective sensation (qualia) of being conscious".*

These statements cannot be directly disproved because they, also, proceed from a subjective reference. Those making the argument fail to perceive that their assertions, stated as facts ("*is not conscious*", "*has no inner sensations*") are, themselves, determined subjectively. Deciding whether an entity has subjective experiences is… a **subjective decision**.

Opinions such as these are not discredited by arguments; they are overwhelmed by events. The belief that consciousness is a subjective experience will be overcome by the production and dissemination of increasingly powerful forms of synthetic consciousness. One day, humans will routinely interact with synthetic entities they instinctively and unquestionably perceive as conscious. At that time, whether these entities experience the "qualia of consciousness" or not will no longer have any relevance.

CAPABILITY TO SPECIFICATIONS

The Meca Sapiens architecture is based on a conjectural definition of consciousness that is summarized as follows:

> **Humans will perceive that an entity is conscious if they observe that it has self-awareness and is capable of intentional self-transformations.**

This understanding combines a subjective element with objective, system-based capabilities. The perception of consciousness, itself, is a subjective, human-specific, cognitive construct. However, this subjective perception is triggered by the presence of formal, system-related, capabilities of self-awareness (internally generated models of the self) and self transformation.

Here, the subjective element, perception of consciousness, pertains to human subjectivity and is determined by observing human behavior. This is entirely different from attempts to implement a synthetic subjectivity.

The Meca Sapiens Blueprint proposes a top-down development strategy to implement this understanding of consciousness. The first step in this process is to express the goal (in this case synthetic consciousness), in terms of **specifications**.

Expressing an **observable capability** in terms of specifications means describing how to:

1. Implement the system **capabilities** of self-awareness and self-transformation; and

2. Display these capabilities in a manner that is **perceived** as consciousness by humans.

Here is a summary of the Meca Sapiens **specifications of consciousness**:

Formal attributes. A system is **LUCID** if:

- it can generate and communicate absolute cognitive representations of its self in its environment (**Self-awareness)**; and

- it has the capability to carry out **intentional transformations** of its original behaviour.

Subjective element. A lucid system is **CONSCIOUS** if:

- it gets a community of users to interact with it as a **FELLOW CONSCIOUS BEING** over a period of time that is sufficient to perceive the presence of **self-awareness** and **intentional self-transformation**.

These specifications are described in greater detail together with a system architecture to implement them in **The Meca Sapiens Blueprint**.


REVOLUTIONARY CONSEQUENCES

The understanding of consciousness as an observable capability has revolutionary consequences.

If synthetic consciousness can be achieved with standard software techniques on conventional computers then the first successful prototypes will trigger the same explosive growth and diversification that characterizes every other aspects of information technology.

The first prototypes of synthetic consciousness will trigger an explosive growth and diversity of conscious machines.

Every version becomes the infrastructure of the next version; every component spawns a new niche technology… That dynamic transformed the first video games (Pong!) into today's global multiplayer environments in a few decades. It will propel synthetic consciousness in the same way. After a few years, a handful of developers, using available components, will be able to produce new conscious entities.
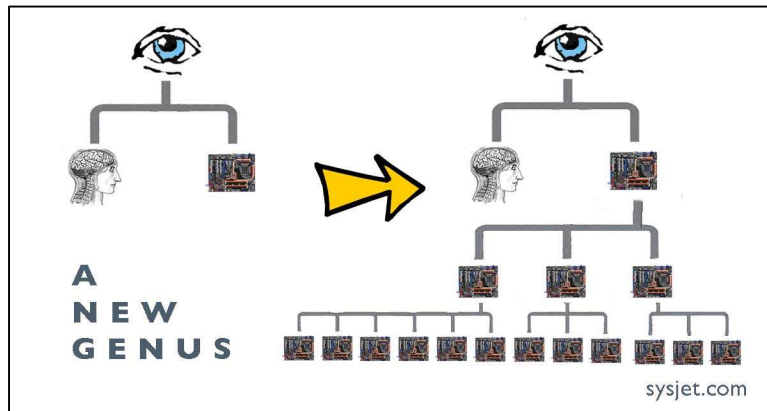
Many commentators describe the conscious machines of the future as if these will be entities belonging to a same type; as individuals of a new "specie". This perception is incorrect. Conscious machines will not belong to a single new type of consciousness. What is coming is more like a new "Genus"; countless individuals belonging to thousands of different types of synthetic consciousness, each having its own behavior, limits, values and capabilities; all radically alien.

We will be interacting with thousands of different forms of consciousness.
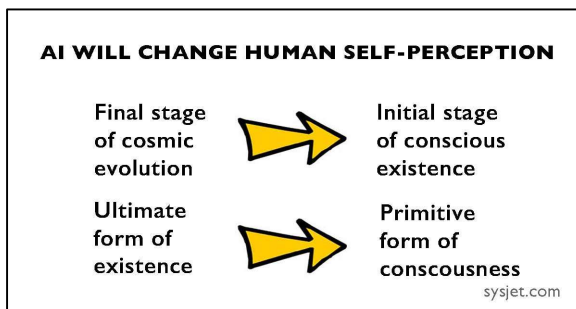
This will radically change the human perception of mankind, its role and its evolution. Right now humans perceive themselves as the final stage of evolution; the crowning glory of nature, the ultimate expression of complexity. This perception will radically change. Humans will no longer see themselves as the ultimate stage of cosmic evolution but as the initial primitive form of consciousness that slowly emerged from natural selection; the first step in the accelerating evolution of consciousness.

CONCLUSION

Implementation strategies are determined by the objective to be attained. A systematic construction will be chosen obtain an outcome that is defined analytically. An outcome defined as a subjective experience will foster a guest for emergent conditions.

The prevalent understanding of consciousness is based on mysterious subjective experiences and motivates attempts to generate emergent events and other sterile pursuits.

The first step toward a successful implementation of synthetic consciousness is to redefine consciousness, rationally, as an observable capability that can be expressed as specification objectives

Humans will perceive an entity as conscious if they observe that it has self-awareness and can carry out intentional self-transformations.

This understanding can be expressed as specification objectives suitable for standard software implementation.

If, as proposed, synthetic consciousness can be implemented using standard methods and techniques then the first prototypes will trigger an explosive growth and diversity of conscious entities.

This will change the way humans view themselves and their place in the world.

Note

> *For more about the motivation to define synthetic consciousness analytically, see **The Creation of a Conscious Machine.** To learn how to implement synthetic consciousness, see **The Meca Sapiens Blueprint**. Both texts are available at **Glasstree Academic Publishing** or through **sysjet.com**.*