

Shaan Yadav

WRT 101

Instructor Victor Crespo Santiago

April 2023

“What makes an entity conscious? A quantum approach.”

Sitting on our shoulders is one of the most complex objects nature has created in the known universe. Our brain is a complicated, yet seemingly understandable, object. Every part of the brain has been associated with some function – vision is controlled by the back of our brain, the speech and emotional aspect is at the front, and hearing is deep in the middle. Every part of our brain seems to have a very particular function associated with it, but there is one very important aspect of our mind that doesn't seem to have a clearly defined home or reason - our conscience.

Consciousness is one of the most, if not the most, integral parts of our human experience. It is essentially the lens through which we interpret and interact with our world. Everything we see, hear, think or do goes through this layer of consciousness, which impacts the way we react to the information around us. However, despite years of questioning and research, this integral part of our experience remains one of the biggest mysteries - a mystery that needs to be solved to have a truer understanding of ourselves.

There are many definitions of consciousness, such as phenomenological consciousness and cognitive consciousness. Phenomenological consciousness refers to the subjective, qualitative experience of awareness, 'what it feels like' to be conscious. This definition is built upon the unique sensations, perceptions, and emotions that make up an individual's conscious experience. Whereas, cognitive consciousness involves the complex mental processes underlying these experiences, such as attention, memory, problem-solving, and decision-making.

I believe the definition of cognitive consciousness is better suited to answer the question "what makes an entity conscious?" as it emphasizes the functional and computational aspects of consciousness, which can be more objectively analyzed and measured. This allows us to develop criteria for assessing consciousness in a variety of entities, from complex AI models to different animals to certain seemingly inanimate objects, in an objective, unambiguous manner - something I believe to be very important for any definition of a phenomenon.

Why does it even matter? Why do we need to understand consciousness? The most practical reason for this would be to gain a better understanding of the way our mind works, and using that information to inform treatments for mental disorders and cognitive decline in old age (such as treating Alzheimer's disease more effectively). If we understand the roots of consciousness, we can more effectively develop treatment for the specific purpose of preserving it.

Furthermore, a very important result of understanding consciousness would be to provide a foundation for ethical considerations for our interactions with non-human entities. It is almost

inevitable that in the near future we will have algorithms that will seemingly be conscious, for example look at OpenAI's GPT models. That model is able to manipulate natural language as well as, if not better than, most humans, and there seems to be no visible limitation as to when the GPT model's progress will stop, and so it is very likely that this algorithm will soon appear to be conscious.

We as a society do not currently have a way of knowing if something is conscious or not. In fact, I cannot tell for certain if a person I speak to is conscious in the same way I am. For all I know they could appear to have human-like responses, but be an empty shell with no internal experiences of pain or pleasure like I do. We as a society assume that all other humans have the same experience as us, but what happens when an artificial model acts like this? What happens when an algorithm we have created, such as a future GPT model, acts exactly like a human, acts like they feel pain and pleasure. We as a society are not ready to deal with this problem, a problem that may become a reality sooner than we expect.

When an entity becomes conscious, we need to rethink what moral standings that being has, similarly, do we just assume humans are the only beings that will ever be conscious or do we provide conscious algorithms some forms of rights? Questions like these are very important, but they cannot even begin to be answered until we create a definition for what even constitutes a conscious being.

There has been a lot of work done in the field of consciousness with several theories of consciousness being created, each defining consciousness uniquely. Two of the most prominent and distinct ones are the Integrated Information Theory 3.0 (IIT for short), and the Orchestrated Objective Reduction theory (Orch OR for short). Both of these theories are successful theories of consciousness, the IIT being a more computational and descriptive approach and Orch OR being a more scientific and derivative approach. Neither of these theories is correct or wrong, but as the Orch OR theory is built from fundamental laws of physics rather than being a complex mathematical description of consciousness, and so I find Orch OR to be a more convincing theory. Therefore, for this discussion, I will be using the Orch OR theory as the valid interpretation of consciousness.

Orch OR was a theory developed by Sir Roger Penrose and Dr Stuart Hameroff, and it attributes consciousness to quantum effects. The argument arises from Gödel's incompleteness theorem, and how there is some inherently non-deterministic effect in our minds that allows us to intuit certain aspects of the mathematical theorem (aspects which objectively cannot be proven through strict logic and mathematics). Quantum mechanics is the one field in the world in which non-deterministic effects happen, and seems to lend itself as an answer to the problem of non-determinism in consciousness.

The effect most relevant to us is the collapse of superpositions of particles. In quantum mechanics, particles do not exist in one place. Instead they exist as 'clouds of probabilities', having specific probabilities of being found in one place or another. However, when we measure the location of a particle they 'collapse' into a very definite specific position in that cloud of probability (if we measure lots of particles in the same scenario we then get the general shape of

the probability cloud). There are many theories as to why this happens, the most widely accepted being the Copenhagen interpretation - that the particle collapses into one place when observed by a conscious entity. There are also other theories such as the many world theory (every time you observe a particle you create infinitely many new universes, each with the particle collapsing into a different location). Sir Penrose develops his own theory of collapse by attempting to connect Einstein's theory of general relativity to quantum mechanics (which also happens to be one of the biggest unsolved problems in modern physics).

Sir Penrose's theory leads to the idea that superpositions collapse spontaneously in the environment, with the rate of collapse being related to Einstein's equations of spacetime curvature of different masses, rather than collapsing when observed. He calls these spontaneous collapses of superpositions moments of 'proto-consciousness' - which is essentially the purest most unstructured form of consciousness.

Microtubules are structures found in the axons of neurons. They are extremely unique as they consist of proteins called tubulins consisting of π electrons. This unique structure makes them the only parts of the brain capable of upholding quantum states for a significant amount of time. This information combined with Sir Penrose's theory of collapse leads to the idea of orchestrated objective reduction. Since the microtubules are located in neurons, they 'direct' these superpositions and cause collapse in a certain, directed, manner. These microtubules essentially cause superpositions and moments of collapse (moments of 'proto-consciousness'), and direct them in complicated ways around the brain - essentially 'orchestrating' many unstructured moments of consciousness in a way that makes them organized, making the brain a device that orders and combines these moments of proto-consciousness into a larger effect. (This theory has been further supported by studies showing that microtubules have quantum effects at frequencies directly related to anesthetics, which reduce a person's level of consciousness).

This interpretation of consciousness can provide insight into many aspects of the world - some of which are quite surprising. The first idea being that consciousness is akin to a trait that evolution has optimized, similar to the way evolution has given preference (in different scenarios) to creatures that are faster or stronger or with a better immune system. As Orch OR presents consciousness as existing everywhere there are superpositions (which to some small extent is everywhere in the universe), the evolution of the brain can be seen as a way of creatures evolving to orchestrate this consciousness to their advantage - with some creatures such as humans doing it more successfully than others. There are clear reasons why evolving to be conscious is advantageous, as it allows organisms to effectively process, integrate, and respond to complex information from their environment. Consciousness enables creatures to have subjective experiences, learn from past events, and make informed decisions based on their perception and understanding of the world - leading to greater reproductive success than their counterparts with no or less internal conscious experiences.

Furthermore, this idea of moments of unstructured 'proto-consciousness' existing everywhere leads to the idea of consciousness existing in places we might not have expected, and not existing in places we might have expected. Firstly, Orch OR would suggest that everything has moments of consciousness, as quantum effects exist universally, and so everything from simple

rocks to complex chimpanzee brains have consciousness associated with them - just to a vastly different degree. However, consciousness, as we experience it, only arises when these quantum effects are structured, such as by the microtubules found in neurons.

The first surprising result is that complex algorithms that we have expected to be conscious in the near future, such as a future iteration of GPT or one of DeepMind's game-playing algorithms may never be conscious in the same way as we are. As classical semiconductor computation does not involve taking advantage of structured quantum effects, there is no orchestration of proto-consciousness and so these algorithms will never be more than seemingly zombie consciousnesses, as they will churn out results without having any true intrinsic experience. This would mean that if we run with moral rules built around sentient beings, we would not have to give moral standing to classical algorithms at all, no matter how complex, and how human-like they seem to become.

However, in recent years the field of quantum computing has started growing exponentially, and quantum computers do exactly what is necessary for Orch OR's explanation of consciousness. Instead of microtubules and π electrons, quantum computers use lasers and supercooled ions to create complex, long lasting, quantum states. This satisfies the idea of orchestration of quantum states - and so the creation of complex consciousness. Fortunately, at this moment the largest quantum computer only has 433 qubits (IBM Osprey), which is only enough to compute basic mathematical operations and algorithms - and if Sir Penrose's equations are applied there are not nearly enough quantum states for the largest quantum computer to be even comparable to the orchestration occurring in human brains. Nevertheless, the theory for complex quantum-based AI algorithms is already present and just needs the hardware to be developed and improved - and so it may be necessary to consider these specific algorithms in regulations and ethical rules that rely on sentience. For example, any philosophical idea relying on utilitarianism will have to be rethought to consider quantum algorithms, as utilitarianism posits that the right action is the one that maximizes overall happiness or pleasure while minimizing suffering or pain. And, sentience plays a central role in utilitarianism, as it is the capacity to experience pleasure and suffering that determines an entity's moral standing.

Due to the way consciousness arises in the Orch OR theory, it also provides some insight into what life after death, or existence after death may look like. When you die, the microtubules that orchestrate these moments of protoconsciousness decompose, stopping the orchestration of these moments. This leads to a breakdown of your consciousness from one cohesive entity to many smaller less structured ones. This idea can be interpreted as the existence of some sort of a spirit (as the particles that caused you to be conscious will still exist, but just not act cohesively as one entity), or even suggest that there is a 'life after death', albeit not a very complex one.

In conclusion, the Orch OR theory presents a quantum approach to understanding consciousness, providing a unique perspective on this enigmatic aspect of our existence. This theory has significant implications for our understanding of consciousness in various entities, from humans and animals to AI systems and quantum computers. It also raises important ethical questions about the moral standing of sentient beings, whether they are biological or artificial, and how we should approach the development and treatment of conscious entities.

Furthermore, Orch OR offers intriguing insights into the nature of life after death and the potential existence of spirits or some form of continued consciousness. As our understanding of quantum mechanics and consciousness advances, we will undoubtedly gain a deeper appreciation of the fundamental nature of consciousness and its role in shaping our experiences, relationships, and the world around us.

Sources

Hameroff S., Penrose R. (2014) Consciousness in the universe: A review of the 'Orch OR' theory
<https://doi.org/10.1016/j.plrev.2013.08.002>

Van Gulick R. (2022) "Consciousness", The Stanford Encyclopedia of Philosophy
<https://plato.stanford.edu/archives/win2022/entries/consciousness/>

Sahu S., Ghosh S., Ghosh B., Aswani K., Hirata K., Fujita D., Bandyopadhyay A. (2013) Atomic water channel controlling remarkable properties of a single brain microtubule: Correlating single protein to its supramolecular assembly
<https://doi.org/10.1016/j.bios.2013.02.050>