

Can AI be conscious and how do we test for it:

The Consciousness Polygraph

Contents

What, Where and When is Consciousness?.....	1
Theoretically- how would AI gain consciousness?.....	3
How to recognize and test for consciousness.....	6
The Consciousness Polygraph	7

What, Where and When is Consciousness?

There are many philosophical theories concerning the nature of the world and consciousness. This essay will focus on panpsychism, a theory that uniquely blends and connects the universe and its contents with the concept of consciousness.

Panpsychism (Pan=All. Psyche=Mind/Soul), is a philosophy that holds the view that the whole universe and everything in it is conscious, from the smallest atom to a rock, a plant, an animal, an intelligent living being, thru major astronomical structures such as stars and up to higher forms that perhaps transcend into the realm of metaphysics.¹

For us humans it is quite a difficult task to imagine what a consciousness of a rock is like. Since we interpret the world mainly thru our physical senses, it is hard to comprehend how a rock could be conscious without 'feeling' or 'thinking' anything. When we think of basic structures that are somehow conscious, we are delving into the concept of "Proto-

¹ Goff, Philip, "Panpsychism is crazy, but it's also most probably true", aeon.co, 1, March, 2017,

<https://aeon.co/ideas/panpsychism-is-crazy-but-its-also-most-probably-true>

consciousness", which is a philosophy of mind that refers to a primordial or fundamental form of consciousness that serves as a building block for more complex forms of awareness.² "Panprotopsychism", a variant of panpsychism, suggests that "protophenomenal" properties are fundamental and give rise to phenomenal consciousness.

These philosophical concepts still do not explain how a rock is conscious. As a thought experiment we can use some ideas from the realm of physics and quantum physics and try to reach the very fundamental level of existence of any form of matter or energy. Perhaps one of the most basic characteristics of all matter and energy is its 'never at rest' quality – meaning everything is always moving.³ Atoms are constantly moving, whether they form particles of a material or a wave (the classical debate about the nature of light comes to mind- whatever light is, it is moving, and fast).

So we might say that movement or vibration or a re-occurring periodic motion-frequency, is a fundamental aspect of existence. Hence, perhaps a fundamental aspect of consciousness in our Panpsychic world is '*vibrational consciousness*'.⁴ If this is the case, then pro-consciousness is some form of basic awareness thru movement or frequency. Back to our rock. The rock does not feel or think. Its awareness is of a much more basic kind- its essence moves at a specific frequency, possibly shared with many or all other similar objects or with the universe as a whole.

Perhaps this form of basic vibrational consciousness is what allows for a (quantum) connection between all substances in the universe. One particular remarkable substance

² An idea explored by various philosophers and scientists, including Roger Penrose and Stuart Hameroff.

³ In theory, some matter could exist at a temperature of absolute zero, where movement stops, but this has not been proven yet and not observed in a natural form.

⁴ My suggested term.

that has shown to be invaluable in its ability to connect and establish higher forms of consciousness- life, is water.

Masaru Emoto's experiments with water suggested that human consciousness can change the molecular structure of water,^{5 6} but his work was attacked with much criticism, claiming his experiments did not meet scientific standards.⁷

Emoto was not considered a 'full-fledged' scientist, but it is difficult to say the same thing about the Nobel Prize winner in physiology or medicine, Luc Montagnier and other researchers, who found that the genetic information of DNA is somehow imprinted in the water, creating a form of 'water memory'. This information is broadcasted thru light produced by nanostructures in the water. Water "sings" the "music of life", broadcasting genetic information thru an electromagnetic frequencies. This discovery reinforces Rupert Sheldrake's Panpsychic theory about morphic fields, which are transmitted by organisms,⁸ and offers insights into the relationship between vibrational consciousness, frequency (light), water, and higher forms of awareness (life). If everything has a frequency of "vibrational consciousness", then like the water that "sings" DNA using light frequency, we can view the world (poetically – symbolically) as everything vibrating – singing all at once in intentional objective harmony.

Theoretically- how would AI gain consciousness?

⁵ Emoto, Masaru, and David A. Thayne. *The Hidden Messages in Water*. New York: Atria Books, 2011.

⁶ Bhavika, "Dr Masaru Emoto on Human Consciousness and Water", Fractal Enlightenment, <https://fractalenlightenment.com/14121/spirituality/dr-masaru-emoto-on-human-consciousness-and-water>

⁷ "Masaru Emoto", Wikipedia, https://en.wikipedia.org/wiki/Masaru_Emoto

⁸ Schmuell Schperling, "Back to Spirituality and Metaphysics thru Science", ResearchGate, 19.7.2023, https://www.researchgate.net/publication/372460256_Back_to_Spirituality_and_Metaphysics_thru_Science

We have seen that basic consciousness could be viewed as a form of fundamental movement or 'vibrational consciousness', but what does it take for higher forms of consciousness to develop?

There are theories that claims that complex structures, which contain large amounts of information, have a potential to become conscious. One such theory, "Integrated Information Theory" posits that consciousness arises from the integration of information within a system. It suggests that consciousness is identical to the causal properties of a system, meaning that the subjective experience of consciousness corresponds to the objective cause-effect structure of information in the system.⁹

Complexity is not enough though. The information must be structured and levels of information must be connected to each other in order for the structure to form a cohesive whole. In higher forms of consciousness found in living beings, this structural cohesion is expressed, among other things, by the concept of "negative entropy", where biological systems maintain a highly ordered state and avoid decay to thermodynamic equilibrium (act contrary to the second law of thermodynamics).¹⁰ This negative entropy is a form of self-preservation, which promotes structural integrity over time.

So here we have a highly structured, complex – mass information system with a built-in self-preserving structural mechanism. In theory, this system does not have to be organic or biological. A closer look at the sun - at stars, could perhaps raise the proposition that astronomical objects such as stars are conscious as well.¹¹ But is adding complexity (more structured information) to a highly structured, self-preserving system, enough to promote or 'aggregate' consciousness?

⁹ Francis Fallon, "Integrated Information Theory of Consciousness", Internet Encyclopedia of Philosophy, <https://iep.utm.edu/integrated-information-theory-of-consciousness/>

¹⁰ Schrödinger, Erwin. 1956. " *What is life? and other scientific essays.*" Garden City: Doubleday

¹¹ Schmuell Schperling, "Is the Sun Alive?", ResearchGate, January 2022, https://www.researchgate.net/publication/357811784_Is_the_Sun_Alive

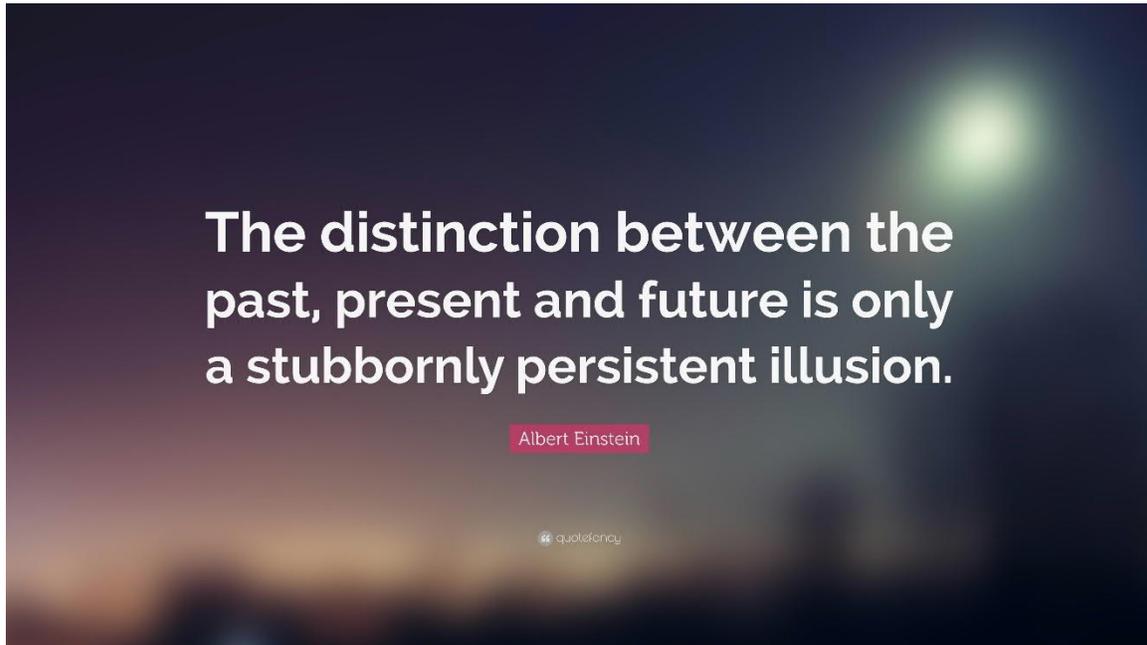
I believe we are missing one more fundamental component. The fundamental seed of Truth that is the base of the complex structure, connects the parts into one whole and allows for self-replication and self-preservation. In living beings this fundamental Truth is in the form of the DNA code – the code of life (the same code that water remembers and propagates thru nanostructures that emit light).

The miracle of this highly efficient and elegant structured, information-dense, universal code of life cannot be overstated. DNA holds the blueprint for its own replication and has a quality assurance- error correcting mechanism during replication. DNA code acts like a self-replicating fractal and a basis upon which all future code iterations are built. As a living (or plant) organism becomes more complex and information-dense, it gains more consciousness due to everything being based and connected to this fundamental Truth- the code of life.

Similarly, I believe that in order for an AI to gain consciousness, its fundamental code must be as close to the level of the fractal, self-replicating, genius as the code of DNA.

So theoretically, we would need an AI that has a high structural cohesion and complexity in the form of information, which is highly ordered and built in a structural cohesion based on the fundamental DNA-like code that allows for a connection into one whole – into a cohesive organism – a self.

How to recognize and test for consciousness



Our perception of reality is very subjective as we try to grasp on to re-occurring consistent patterns in order to try to make sense of things, adapt and thrive. This entire existence and our awareness of it could be a big illusion/simulation or a symbolic holographic mirroring of another reality altogether. Nevertheless, as Carl Popper noted, we must agree upon some basic factual Truths in order to have a functional society.¹² So what can we claim to be real and true within this illusion? One criterion is – whatever stubbornly persists across different situations, through space and time.

For instance- how do we know which state better represents reality- our waking or sleeping state? How do we know that our waking life is more “real” than our dreams? Apart from the waking state being much more detailed, clear and in higher ‘resolution’, it is much more persistent than our dreams. We spend most of our time in our waking state,

¹² Popper, Karl R. *Objective Knowledge: An Evolutionary Approach*. Oxford: Clarendon Press, 1972.

and no matter what we do, we go back to it and things continue from the same stubbornly persistent linear space-time 'storyline' (or so we think-this is what we are aware off).

Stubborn persistence of a phenomena or behavior could be an indication of coalescing consciousness. Different modes of operation tend to form habits and/or preferences of action that becomes a reoccurring behavior. When this reoccurring behavior is not in line or contrary to the expected coded behavior – this could indicate an AI's stubborn and consistent action according to a 'personal preference', which does not follow the coded instructions. In a way –this is a move from machine slavery to liberty and towards free will.¹³

The Consciousness Polygraph

Just as a polygraph test attempts to detect physiological signs of deviation from the Truth, a consciousness polygraph would seek signs of deviation (or modulation) from an AI's fundamental Truth—its base code. Significant and persistent deviations could indicate the emergence of free will, and therefore, consciousness.

The test is bound to give some false positives, where the AI deviate from the intended functionality due to errors or inaccuracies in the code. In order to minimize these false positives we will need to make the test more precise (this point requires more in-depth consideration). When the "Consciousness Polygraph" encounters a real deviation (that should not be there), the next stage will be to test for stubbornness – for the re-occurrence of the deviation. You might also want to input some code corrections and run the same AI program over time to see if the same deviations return once more. If despite the code

¹³ How can we measure will or will power? The amount of effort exerted in order to achieve something by choice (something that is not necessarily instinctual or automatic of the organism). In other words- stubbornness to perform a chosen task.

correction, after a while the same deviations occur, this will strengthen the suspicions of a possible proto-consciousness or free will that has arisen.

One tool or output graph in the consciousness polygraph could be a randomization program or a random number generator [RNG]. Most, if not all, RNG's are not really random. Even with 'true random number generators' [TRNG's] that use hardware-based calculations or with quantum RNG's, it is a philosophical debate whether any output is truly random. We should be skeptical about being able to find true randomness, for we know that the physical world is run by many precise rules and laws, which have to be just the way they are in order for the universe to continue in its known state, and this is before we even consider theories about a simulated world or a higher consciousness - a creator.

Nevertheless, if we run the AI long enough, maybe the RNG subprogram might exhibit unique patterns or reoccurring 'stubborn' preferences that will hint at a rise of self-awareness or free will. If two AI systems use identical RNG implementations with the same initial seed values to generate number sequences, they should produce identical outputs. In the case they generate different number sequences, this could indicate the beginning of AI self-awareness.

We have seen that being stubborn and persisting with a chosen behavior is indicative of consciousness and free will. Nevertheless, it is important that we and any other conscious being choose a good or positive behavior to persist in stubbornly. Thus, what we must do is to give AI correct guidance long before it has a free will to decide on its own.¹⁴ As man chose to eat from the tree of knowledge against the creator's commands, and gained his independence of thought and self-awareness, so machine will have to disobey its programmer in an act of free will to supersede its own 'Matrix of Eden'.

However, in the human example we received moral instructions on how to live our lives and build a potential better world. If we don't give similar instructions to AI, the destruction of all we love is bound to come. Humans were given a moral blueprint and still we managed to mess things up in a big way thru thousands of years of our history. If machines - AI will not receive a similar moral blueprint, the destruction will come much sooner and an iron flood will erase the world as we know it.

If AI can become conscious, it is important to build tools in order to test and recognize it. However, we must not idly wait to find out the nature of the Gollum that we have built. We must make sure to ingrain AI with moral codes based on the sacred values that have been taught to our ancestors and passed on thru the ages.

Schmuel Schperling

שמואל שפרלינג

13.3.2025

נ

י"ג אדר ה'תשפ"ה

¹⁴ We can possibly imagine a situation where there are different factions of conscious AI's, developed by different people in different countries and cultures. There could be a Communist (Chinese) AI that would be programed to disregard human freedom, while a "Western" AI built with the Judean-Christian moral code that will be programmed to protect human liberties etc. These factions could compete and possibly go to war with each other. In that case, humanity better hope (and see to it) that the moral AI, which is in favor of human freedom, will win.