Physics and Consciousness

Han Liu

Advisor:
Professor Izabela RC Stroe

WPI

This project report is submitted in partial fulfillment of the degree requirements of Worcester Polytechnic Institute. The views and opinions expressed herein are those of the authors and do not necessarily reflect the positions or opinions of Worcester Polytechnic Institute.

Abstract

This MQP - Physics and Consciousness presents several existing theories of consciousness. Based on these theories and other scientific findings, a new preliminary theory about the physical foundation of consciousness is proposed and discussed.

Table of Contents

1. Introduction

Consciousness has long been an intriguing and mysterious problem for humans, an understanding of physical foundation of consciousness and intelligence could be the most significant achievement in the 21st century.

Roger Penrose and Stuart Hameroff's 'Orch OR' of consciousness suggests that the microtubules within brain neurons is the physical location of consciousness. When electron clouds of microtubule protein molecules undergoing the reduction to classic states from a quantum one caused by a gravitational energy discrepancy of superposition states in space-time evolution (Objective Reduction), a proto-qualia is 'generated'. The 'orchestrasted' massive 'ORs' among neuron cells with particular frequencies then result in the frames of consciousness in the brain. This theory is by far the most specific one concerning the exact physical process of consciousness, there was also a range of evidences showing that neuron cells do not behave totally classic. However, they didn't go further into the properties of qualia as the physical assumption is by itself the 'Hard Problem' for quantum mechanics.

Giulio Tononi proposed 'The Integrated Information Theory of Consciousness' as a hypothesis of the functionality part of Consciousness. It assumed that the dynamic 'integrated information' of a system gives the degree of consciousness, with five axioms: intrinsic existence (consciousness is intrinsic and real), composition (consciousness is structured), information (consciousness is specific and differing from each other), integration (consciousness is unified and irreducible to non-interdependent subsets) and exclusion (consciousness is definite). IIT avoids the 'Hard Problem' and tried to construct a causal system unprecedented and focused on what kind of systematic properties should fit in with and eligible for consciousness (Minimum Information Partition), thus establishing criteria for judging the existence, magnitude and specific states of consciousness of a system. The theory is mostly tenable empirically, yet the mathematical model is unclear and flawed.

The Turing machine-based AI system which is limited to a fixed set of axiomatic rules will not be able to find rules outside the system, while real intelligence system like human brain has not such limitation for the inputs of real randomness from intrinsically random physical process. The evolution suggested a tuning process of consciousness toward a better cooperation with functional intelligence.

To avoid the problem of which immense hidden properties must be planted in certain physical mechanism for an explanation of consciousness, the theory of background consciousness field is proposed. Instead of creating proto-consciousness by themselves, physical properties are assumed to be filters of qualia from the background field.

2. The 'Orch OR' Theory

2.1 Introduction

Consciousness is one of the biggest mysterious phenomena that are perceived yet not understand. It is the subject experience originated of both the inside and outside world. In general, there are three general views of consciousness.

The first type of view take consciousness as an non independent element that is attached to certain mechanism of the physical world. It emerges as a result of biological evolution in billions of years and the benefits of consciousness can therefore be assumed as something necessary for better adaptation to the environment. It is unknown to what complexity of life form or particular structure that perceivable consciousness is available for making a difference.

The second type of view thinks that consciousness permeates in the universe: It is everywhere and unlike matter, which is confined to strict laws of physics, consciousness has properties distinct from the physical world and even beyond the realm of science (e.g. impossible to build a precise mathematical description that concludes and predicts every aspects of consciousness). Moreover, some theories claimed that the material world is actually a fake illusion while what actually exists is consciousness itself.

The third type of view assumes that consciousness is the result of exact physical mechanism which is not found yet or has been neglected. The physical laws, undiscovered or not, will be able to fully describe the nature of consciousness. The evolution of consciousness is the temporal and spatial arrangement of physical events (which generates proto-consciousness) so that the combined consciousness is capable of give cognitive functions. It is supposed by some scientists that this process is linked with the collapse of wave function (reduction from quantum state to classical state).

Stuart Hameroff and Roger Penrose (Dennett, 1991) "proposed in the 1990s that the consciousness depends on biologically 'orchestrated' coherent quantum processes in collections of microtubules within brain neurons", "the continuous Schrodinger evolution of each such process terminates in accordance with the specific Diosi-Penrose scheme of 'objective reduction' of the quantum state" (Hameroff, 2014). "This OR activity is taken to result in moments of conscious awareness and choice." (Hameroff, 2014) They "also introduce a novel suggestion of 'beat frequencies' of faster microtubule vibrations as a possible source of the observed electroencephalographic 'EEG' correlates of consciousness." (Hameroff, 2014)

2.2 Consciousness, Computation and Neuronal Information Processing

There are several unknown properties of consciousness, which retains essential: The intrinsic components of consciousness (qualia) that are still outside the scope of mathematical and physical description; The mechanism which collective and integrate subjective perceptions based on seemingly disparate activities on various locations of a neuron system - the synchronized pace of these activities and what level of physical model (classic electromagnetic theory of quantum mechanics) is needed; The capability that a conscious mind exceeds what a typical (Turing) computer in understanding concepts and creating new paradigms (Hameroff, 2014).

Currently recognized as a series of frames of single but integrated picture of experience/perception, the frame rate is observed most related to gamma EEG, which is synchronized membrane electric activities between 30 to 90 Hz among different regions of the brain (Hameroff, 2014).

Most neuroscientists assumes that it is the complicated synaptic activity on the neuron network that functions as both the basic information processing unit and the location where consciousness (at least the elementary component) comes into being (Hameroff, 2014). According to the 'Hodgkin - Huxley' model, all neurons behave in a classic 'integrated-and-fire' threshold model which resembles classic logic devices (Hameroff, 2014). While this model and claimed that the integration of electric potential and axonal firing give rise to consciousness, Hameroff argued that gamma EEG is on the other hand created by the dendritic and somatic integration potentials (Hameroff, 2014). The active

integration in the Hodgkin-Huxley model is classic and deterministic, while in cortical neurons the threshold at axon initiation segment changed almost every time, which indicates mechanisms other than traditional classic 'integration and fire' model (Hameroff, 2014). In addition, some molecules that affect postsynaptic dendrites can wipe consciousness leaving axonal integration mechanism unimpaired (Hameroff, 2014).

As often omitted in some models and theories, there are gap junctions between dendritic membranes and there are microtubule networks inside both dendrites and axons.The Orch OR theory suggests that the quantum mechanisms on microtubules when signals integrate between dendrite and axons actually decide the result, while the dendritic networks (connected by the gaps) can synchronize in order to organize neuron activity on a much larger scale (Hameroff, 2014). These properties suggests ideal locations for perception and action with the emergence of consciousness (Hameroff, 2014).

Gap junctions make it possible for various connected dendrites to have synchronized local field potentials thus to increase the scale of inputs and the computation power. While the result - axonal firing potential is further decided by the microtubules inside the dendrites, so the threshold would be different with similar inputs.

Cytoskeleton is the basic interior 'scaffold' of eukaryotic cells, which is composed of protein network of microtubules microtubule-associated proteins (MAPs), actin, and intermediate filaments (Tuszynski, 1995, as cited in Hameroff, 2014). In OR theory, cognitive functions can be achieved along on cytoskeletons even without synaptic activity (Hameroff, 2014). "Microtubules ('MTs') are cylindrical polymers 25 nanometers in diameter, and of variable length, from a few hundred nanometers to meters in long nerve axons (Hameroff, 2014)." "MTs self-assemble from peanut-shaped 'tubulin' proteins, each tubulin being a dimer composed of alpha and beta monomers, with a dipole giving MTs ferroelectric properties (Hameroff, 2014)." "Tubulins are usually arranged in 13 longitudinal protofilaments whose lateral connections result in two types of hexagonal lattices (A-lattice and B-lattice), the protofilaments being shifted in relation to their neighbors, slightly differently in each direction, resulting in differing relationships between each tubulin and its six nearest neighbors (Hameroff, 2014)." Helical pathways exist in the structure repetitively (e.g. in some fixed number of tubulin monomers along the microtubule polymer). It is worth noting that the Tau proteins (if bound to MT) is essentially in synaptic plasticity and encoding information as their locations on Microtubules are also the traffic signals. People

often argue that it is hard to find a place for quantum mechanics in the hot, damp brain tissue, yet microtubules, which would have very little structural change (as mature neuron cells do not split), would provide such stable locations. In addition, in each neuron there are roughly 10^9 tubulins - note that this far exceeds the number of synapses, which is usually 10^3 per neuron (Hameroff, 2014).

Hameroff and Watt suggested that the structure and physical properties of tubulin - the dipole directions as well as the shape of the molecules are the storage location for information, while the Microtubule lattices functions as boolean switching matrices (Hameroff, 2014).

Typically there are different time scales of structural change on proteins (e.g. 10^{-6} s to 10^{-11} s), which is an energy consuming process while also creates heat (and lack of evidence in terms of massive scale information processing). Previous OR theory assumed that structural change would be the computation process, however in late edition only the dipole directions are used to store and convey information (Hameroff, 2014). The information from outside and be 'entered' on the microtubules by chemical receptors (e.g. MAP2 and CaMKII) - they causes change on lattices and the states of the dipoles are affected. According to latest OR, although there are various states involved, the modeling of this computation process focuses on two alternative states (Hameroff, 2014). The repetitive pathways where dipoles are aligned along would be a potential functioning place for information processing (Hameroff, 2014).

The roles that microtubules played in the latest OR were supposed to be: (1) MT processing during dendritic–somatic integration can influence axonal firings to implement behavior (Hameroff, 2014). (2) MT processes may directly result in conscious awareness (Hameroff, 2014). (3) MT processes can regulate synaptic plasticity (Hameroff, 2014). (4) Tubulin states can encode binding sites not only for tau, but also structural MAPs determining cytoskeletal scaffolding and thus directly regulate neuronal structure, differentiation and synaptic formation (Hameroff, 2014). (5) MT information processing may be directly related to activities at larger scale levels of neurons and neuronal networks (Hameroff, 2014).

The computational capability of a model based on tubulin dipole states is gargantuan, Hameroff (2014) assumed a scenario of 10^9 logic switches (on MTs) working at 10^7 Hz on one single neuron, which renders 10^{16} operations in one second. Traditional estimation based

on synaptic potential firings has 10^5 operations (10^3 synapses working at 100 Hz) per neuron and 10^{11} neurons in the whole brain. This means that tubulin computation is 10^{11} times more powerful than classic model and one neuron is as capable as entire 'classic' brain. However, as modern cognitive science has pointed out, the frame rate of consciousness moments is not likely to surpass 100 Hz and visual information is the major content of that 'picture', and human eyes has only limited definition (the perceived definition is actually even much lower than the physical limit of the eye). To support such level of information processing, the 10^{16} brain is more than capable with modern algorithms and the astounding 10^{27} quantum brain is questionable as the efficiency of 'natural' algorithms in the brain will be laughably low - or at least a sign that most of these operations were not purely for data processing.

The tubulin "is composed of a heterogeneous group of amino acid residues connected to peptide backbones" (Hameroff, 2014). Anesthetic gas molecule can affect consciousness by interacting (e.g. london force) with intra-protein hydrophobic regions on tubulins (a very strong indication that tubulins are related to consciousness). In former OR theory, tubulin dipoles were thought to be driven by London-force, while it is assume in the lastest theory that the electron spin or nuclear spin, which is connected with magnetic dipoles and have a rather longer life, is the framework of the computation (Hameroff, 2014). The spin is one of the earliest known Quantum mechanism, it is by nature a good medium for quantum computation. The flip of the spin may be a reason for the alternating currents on microtubules and it has been found that the transfer of spin on the aromatic rings is strengthened when the environment is hotter (Ouyang, 2003, as cited in Hameroff, 2014).

The 'up' and 'down' state of the spin (when the magnetic moment is aligned or anti-aligned) can be directionally chosen as basics states for quantum computation, while it naturally allowing properties like quantum superposition (Hameroff, 2014). The OR speculated that "chains of correlated ('up-up-up', 'down-down-down') or possibly anti-correlated ('down-up-down', 'up-down-up') spin along lattice pathways in microtubules might provide biologically plausible ways of propagating quantum bit pairs (qubits) along the pathways" (Hameroff, 2014). "In pathways, periodic spin-flip or spin-precession processes (either electric or magnetic) might occur and could be correlated with alternating currents in microtubules at specific frequencies" (Hameroff, 2014).

The best experimental finding related to the spin-flip and alternating current phenomenon is that only alternating currents with specifically selected frequencies in GHz, MHz and KHz would invoke resonances on microtubules (Sahu, 2013, as cited in Hameroff, 2014). As quantum spin flip does not allow continuous resonance distribution in any pathway.

Furthermore, dipole shifts can cause femtometer level of change in location of nucleus (as suggested by OR) (Hameroff, 2014). Tiny as it is, the displacement plays a key role in OR.

2.3 Quantum Physics and Consciousness, Orch OR

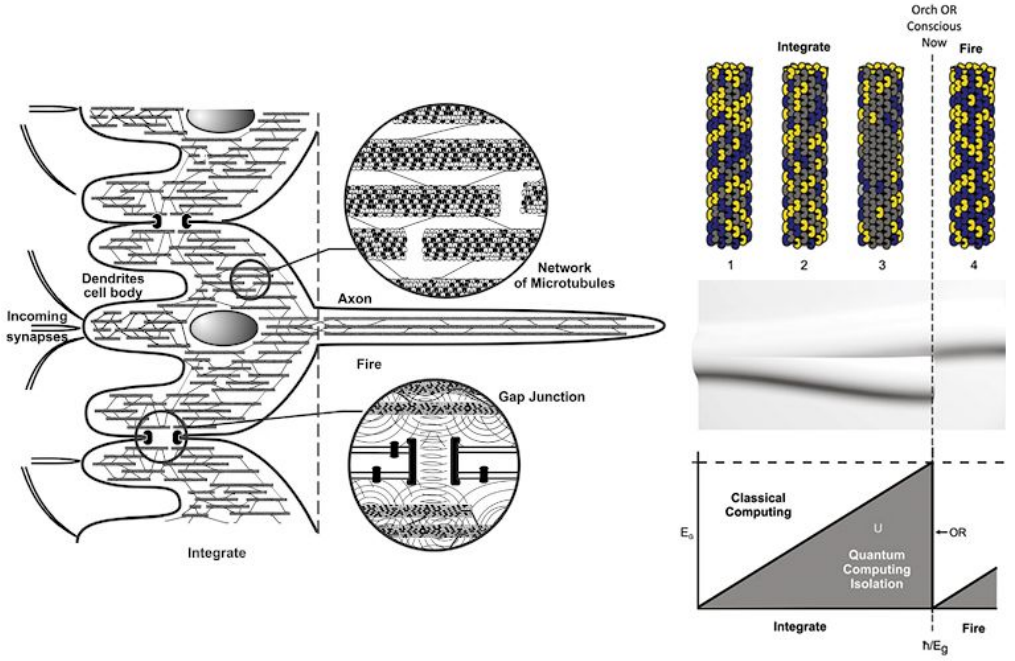

Fig. 1. Left: Dendrites, axons, gap junctions and internal network of microtubules in neurons. Right: A conscious event suggested by the OR theory where the integrated superposed states collapse to one state (up) when OR threshold (down) is reached (as space-time curvatures (in the middle picture) reaches that threshold). (Hameroff, 2014)

The 'strong artificial intelligence' is strongly challenged by the view in *The Emperor's New Mind* (Penrose, 1989), where Penrose showed that the Gödel's theorem doesn't allow all mental processes to be computational and therefore the consciousness is at least in part non-computable and new physical theories are needed to fill the blank (Hameroff, 2014).

The DP (Diósi–Penrose) proposal provides a physical explanation that involves measurement of space-time curvature and gives out a threshold when and where quantum

superposed states would likely to return single ones. In each of this OR event, a moment of 'proto-conscious experience' is created and perceived (Hameroff, 2014).

The nature of quantum mechanics gives the relation of energy and the oscillation frequency (which can only select discrete values): E = hv (Hameroff, 2014).. Quantum states can superpose with other states like waves, yet in the macro world this is not observed (the measurement problem) (Hameroff, 2014). In addition, it has been observed that entangled particles are still 'integrated' even with spatial distances (which is derived from quantum mechanics), while this phenomenon is not capable of transmitting real data, because a traditional channel is always needed to verify the entanglement after comparing the 'collapsed' states after measurement in separate locations (the states are unknown before the measurement).

In this stage of OR, it is the measurement problem that is most related to consciousness (Hameroff, 2014). The development of quantum states are determined (deterministically) by the fundamental Schrödinger equation, and this unitary evolution (U) is somehow 'stopped' and 'randomly' reduced to another state - the process is named reduction (R) (Hameroff, 2014)..

In DP proposal, Einstein's general relativity is considered as the key of triggering reduction. The U is seen as real and objective process - quantum-superposed alternatives in space-time (Hameroff, 2014). When the differences of gravitational energy distribution of the alternative states in space-time increase, the superposition will be more likely to collapse. A timescale $\tau \approx h/2\pi E_G$ is given by DP as a 'kind of average time for the state reduction to take place' (Hameroff, 2014). The R process is random by nature, E_G is the 'gravitational self-energy of the difference between the two (stationary) mass distributions of the superposition' (Hameroff, 2014). Roughly speaking, E_G is the energy cost of moving one state from the original center (before evolution) to the current location, while in the gravitational field of the other state (Hameroff, 2014).

As shown in Fig. 1, two alternating states indicate 2 different mass distributions and the bifurcating space-time histories - the space-time curvature keeps increasing with the evolving of the mass distributions (Hameroff, 2014). The detailed level of separation is described ' in terms of a symplectic measure on the space of 4-dimensional metrics', 'the product of the temporal separation T with the spatial separation S that measures the overall degree of

separation', when the separation hit a certain level, the alternating history begin to (in much greater chances) collapse into one of the states (Hameroff, 2014). The difference required for OR is 4-volume Planck measure, which is extremely tiny in space-time: if the time scale in the measure is small then the space difference will be 'relatively large', if the space difference is very small, a rather longer time is required to make it happen (Hameroff, 2014). Hameroff (2014) proposed that a cat (10kg) would need a time scale of 10^{-43} s to reach the threshold and in comparison, a single electron would require thousands of years.The quantity of the space difference is measured by $S \approx E_G$, (while $E_G \approx h/2\pi\tau$). To put it another way, OR happens when time reaches critical value - in average, $\tau \approx h/2\pi E_G$ (Hameroff, 2014).

In OR theory, the state reduction is also the process of the emergence of 'proto-consciousness' - the basic element of more complicated consciousness (subjective experience) (Hameroff, 2014).

The Orch OR is a 'orchestrated' (massively synchronized and contain information for computation) quantum superposition reduction which is isolated from random environment interference (i.e. only decided by the intrinsic measure of $\tau \approx h/2\pi E_G$), and Orch OR will bring 'Orch' 'proto-consciousness' - which at some scale is one 'frame' of consciousness (Hameroff, 2014).

It is not surprising that this process could happen in temperatures far above absolute zero. 'Biology appears to have evolved thermal mechanisms to promote quantum coherence' (Hameroff, 2014). Ouyang and Awschalom (2013) showed that 'quantum spin transfer through phenyl ring π orbital resonance clouds are enhanced at increasingly warm temperatures (Spin flip currents through microtubule pathways)' (Hameroff, 2014). In photosynthesis, it has been found that energy of a photon absorbed will be transferred to other locations through mechanisms (pathways through π electron clouds) that require electron quantum conductance, as similar things were assumed by the OR theory (quantum conductance along the microtubules) (Hameroff, 2014). Furthermore, mechanical vibration (at some frequencies) can enhance quantum conductance in photosynthesis (Sahu, 2013, as cited in Hameroff, 2014), while Hameroff (2013) discovered that 'Low intensity ultrasound (megahertz mechanical vibrations) administered through the skull to the brain modulates electrophysiology, behavior and affect, improves mood in patients suffering from chronic pain'- apparently 'Orch OR' is also enhanced.

Anirban Bandyopadhyay and his colleagues have researched the electronic property of single microtubule with nanotechnology. They found that microtubules can turn from insulators to conductors under the stimuli of AC current in a range of frequencies (GHz, MHz and KHz) (Hameroff, 2014). The induced conductance is also directional, just like the quantum pathways on Microtubules (Hameroff, 2014). In addition, a 25 nanometer wide microtubule has more AC conductance than a 4 nanometer wide counterpart, which suggests that certain synchronized or coherent quantum mechanisms may take place (Hameroff, 2014). Again, this is happening in typical human temperature - while other studies have also shown quantum effects in similar environments (e.g. bird-brain navigation (Gauger, 2011), ion channels (Bernroider, 2005) sense of smell (Turin, 1996), DNA (Rieper, 2011), protein folding (Luo, 2011), and biological water (Reiter, 2011)).

The beat frequencies emerge as states with slightly different energies superpose and OR is introduced (similar to the beat of classic waves). For energy E1 and E2 (E1≈ E2), the 'classic' frequence is given by $|E_1 - E_2|$/h according to DP, with $e^{ia} + e^{ib} = 2\,e^{i(a+b)/2}\,cos\frac{a-b}{2}$ (take a = $-E_1$ t/h and b = $-E_2$ t/h), the chance to find each energy state is $\{1 + cos(a - b)\}/2$ and $\{1 - cos(a - b)\}/2$ respectively (Hameroff, 2014). The other frequency is the quantum oscillation frequency $(E_1 + E_2)$/2h, Hameroff (2014) stated that OR is not just a process that superpositions reduce to a certain state or location, it is also a process that quantum oscillations reduce to classic ones (beat frequency). While $\tau \approx h/E_G$, it will be much larger than quantum oscillation period $2\pi h/(E_1 + E_2)$ and it could be much smaller than beat period $2\pi h/|E_1 - E_2|$ (Hameroff, 2014). That indicates the beat frequencies will remain almost the same while the phase (affected by reduction) would change a lot ($(E_1 + E_2)$/2h is much higher than $|E_1 - E_2|$/h) (Hameroff, 2014).

Orch OR events are suppose to happen with beat frequencies $|E_1 - E_2|$/h (while it is against the fact that consciousness as a whole has a frequency of E/h) (Hameroff, 2014). The time length of superpositions required will safely be anything shorter than $|E_1 - E_2|$/h (e.g. Hameroff provided an example of two frequencies, 10.000000 MHz and 10.000040 MHz for E_1 and E_2, then the beat is only 40 Hz) (Hameroff, 2014).

2.4 Orch OR and Quantum Brain Biology

In terms of quantum computation, as shown in Fig. 1, the microtubules are quantum processors: the U stage conducts quantum bits processing, and the R gives the output - classical bits - while at the same time giving rise to consciousness (it is often assumed that R is related to consciousness, not merely OR theory, however all these attempts have to mix two problems together) (Hameroff, 2014).

In the 'beat frequency' model, Orch OR events happen at very high frequency and compose consciousness frames in beat frequency (Hameroff, 2014). The proto-consciousness cause by single OR is random and not recognizable, it is the orchestrated massive OR that would meet $\tau \approx h/2\pi E_G$ (during which the environment can not interfere the evolving) and create conscious moments (Hameroff, 2014).

As for proto-consciousness, it may be the most elementary property of the most basic space-time structure (in planck scale). Tiny as it is, the energy involved (E_G) is also so small that it is neglectable as compared with other energy consuming processes in life (Hameroff, 2014). In that case, consciousness (or the attempt to create one) has a nature of extremely low energy cost. In addition, E_G is not even comparable to other biological energies. Since the gravity is so weak, there would be a large amount of matter involved so that the time τ will be relatively small for OR (Hameroff, 2014). Microtubules and other structures would (as evolved to) superpose and through some mechanisms linked to the basic space-time structure and give rise to consciousness frames as well as process information and affects neuronal physiology (Hameroff, 2014).

To find a timescale roughly the same with cognitive experience (length or gap of consciousness frames), as indicated from gamma EEG and other brain wave frequencies, τ is ranged from 0.01s to 0.5s (Hameroff, 2014). To consider the magnitude of E_G, there are three levels of mass and separation: the functioning protein, atomic nuclei and nucleons (Hameroff, 2014). If we adopt the carbon nuclei with a length of 2.5 femtometers, τ is then

0.025s and the corresponding frequency is 40Hz, Ec is then given by the gravitational field energy Ec = Gm^2/ a_c (a_c is the carbon nucleus sphere radius) (Hameroff, 2014). According to hameroff (2014), this requires 2×10^{10} coherent self-collapse from superposition on tubulins in 0.025s. That level of synchronization would need only 2000 to 20,000 neuron cells , way below the amount of total neurons in the brain but in accordance with some estimations in neuroscience (Hameroff, 2014). On the contrary, roughly 10^9 neurons are needed in the beat frequency model while τ can be much smaller (Hameroff, 2014).

A main assumption of OR on structural and function of neuron cells is that gap junctions can transmit molecules and electric potential between neighbour cells so that microtubules networks in one neuron can extend and connect to others, making it possible for brain level synchronization (Hameroff, 2014). In this context, both beat frequency model and simple coherent collapse model are plausible. Studies (Fukuda, 2000, as cited in Hameroff, 2014) has shown evidence that gap junctions is essential in the gamma synchrony and related to consciousness (Hameroff, 2014). In addition, Hameroff (2010) developed the 'Conscious pilot' model in which 'syncytial zones of dendritic gamma synchrony move around the brain, regulated by gap junction openings and closings and in turn regulated by microtubules' (Hameroff, 2014). This further explained some functional properties observed and tested (empirically). Particularly for the bistable perceptions illusion (when one recognize different objects/patterns from the same picture of shape), the OR make it possible to reduce to one of the alternating patterns (e.g. face or vase) from superposed states.

3. IIT and Perceptronium Theory

3.1 Integrated Information Theory (IIT)

Giulio Tononi proposed the integrated information theory as an approach for bypassing 'the hard problem' and instead trying to solve a 'pretty hard' question - The basic structural and mathematical requirements of a physical system that would allow consciousness to take place.

IIT assumes that systems are conscious to the exact degree their dynamics encode integrated information (Johnson, 2016). Being 'integrated information', it can not be dissected and simply seen as a permutation of elements (e.g. perception of a red triangle is not the

combination of perception triangle and perception red). A system's qualia is mathematically described given how its parts are interrelated with each other (Johnson, 2016).

There are five axioms of IIT: intrinsic existence, composition, information, integration and exclusion (see Fig.2). The axioms are categorizes so as to fit with all human consciousness. Moving forward, the properties of physical system is discussed and phenomenological algorithms are derived (Johnson, 2016).

Tononi developed several concepts for IIT, and those concepts are not confined to scales (e.g. from nanoscale to neuron groups) (Johnson, 2016). IIT measured the subsets of a deconstructed casual system, then find their 'MIP' (Minimum Information Partition) - the subset that most determines its past and future states (Johnson, 2016). 'MIP' is then reconstructed in 'cause-effect space' (a vector space), the amount of information of the reconstructed 'object' is denoted Φ, an indication of the level of consciousness.

IIT has successfully explained some states of consciousness (with its alteration): e.g., 'integration (as measured by EEG-based perturbation) is relatively high during wakefulness, decreases during slow-wave sleep, and rises during REM sleep (Massimini et al. 2005), and is lower in vegetative coma patients than those that later wake up' (Casali et al. 2013, as cited in Johnson, 2016). Moreover, Φ is a good index of the complexity of a cognitive task that required intelligence - and this Φ links intelligent behavior to consciousness.

According to tononi, other 'MIP's with a non-zero Φ would be related to consciousness malfunctions like dissociative disorders and other psychiatric conditions (Tononi and Albantakis 2014, as cited in Johnson, 2016).

The general picture is that 'MIP's move and compete with each other (through their Φ value), deciding the final consciousness state (tononi assumes that usually the cortical neuron networks with rich interconnectedness would 'win') (Johnson, 2016).

The IIT is actually an attempt to build the topographic map for consciousness and intelligence (although these two are somehow mixed). However there are still lots of flaws: The clarity of definition of Φ is of questioned and there we don't see how the axioms can derive this value (Johnson, 2016). In addition, the axioms are only empirical, it may seem plausible to some degree, yet no reason is given and no physical proof has been found to show that the axioms (only a phenomenological assumption from physics' point of view) are irreplaceable (while someone can easily bring up a similar theory with 4 or 6 distinct axioms).

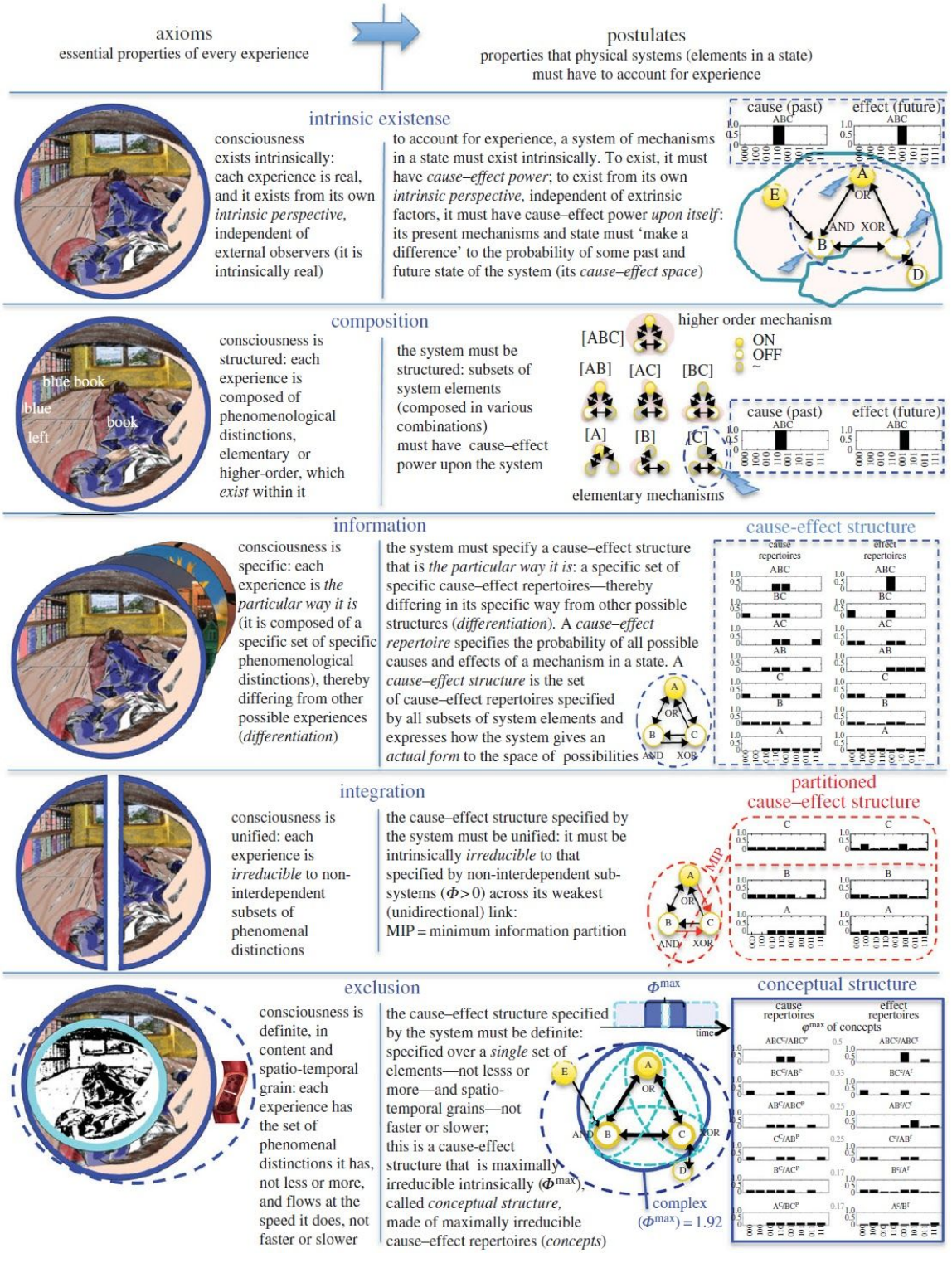

Fig. 2. IIT's axioms and postulates. (Tononi and Koch 2015).

3.2 Perceptronium Theory

Max Tegmark proposed the Perceptronium theory (Tegmark 2015, as cited in Johnson, 2016), a theory that rebuild the IIT with quantum mechanics (Johnson, 2016). Similar to the OR theory, The percetronium theory assumed that consciousness is directly linked to the physical world, and it has to be in harmony with quantum mechanics (Johnson, 2016). The idea is that the combination of energy states in numerous interactions in the brain is a tank for searching for the physical foundations for IIT. Furthermore, the competition of 'MIP's is considered as a variation of the quantum factorization problem (Johnson, 2016). He believed that we need to find those conditions and requirements 'to narrow down what sorts of factorizations of Hilbert Space could support these requirements' (Johnson, 2016). In this way, the mystery of unified consciousness will become a physics problem.

Tegmark has identified six principles: (1) Information principle: 'A conscious system has substantial information storage capacity'; (2) Dynamics principle: 'A conscious system has substantial information processing capacity'; (3) Independence principle: 'A conscious system has substantial independence from the rest of the world'; (4) Integration principle: 'A conscious system cannot consist of nearly independent parts'; (5) Autonomy principle: 'A conscious system has substantial dynamics and independence'; (6) Utility principle: 'An evolved conscious system records mainly information that is useful for it' (Johnson, 2016).

The Perceptronium theory is a good supplement for IIT, with much better mathematics and links to physical mechanisms. However, just as the IIT, it only has a guideline for what the physical system should be like and the theory even confines itself inside the postulates, making it less 'universal' than IIT (Johnson, 2016).

		$\phi^{2.0}$	$\phi^{2.0'}$	$\phi^{2.0''}$	$\phi^{3.0}$	ϕ^{M}	ϕ^{B}	$\phi^{M}_{kk'}$	ϕ^{oakk}	ϕ^{opkk}	ϕ^{otsk}	ϕ^{ofuk}	ϕ^{nask}	ϕ^{mask}	ϕ^{xfkk}
Major	Always **non-negative**	y	y	y	y	y	**N**	y	y	y	y	y	y	y	y
	Always **finite** even for ∞-dimensional system	**N**	y	y	**N**	y	y	y	y	y	y	y	**N**	y	y
	Vanishes for **deterministic** system (drawback)	n	n	n	n	n	n	n	n	n	n	n	n	n	**Y**
	Vanishes for **separable** system	y	y	y	y	y	**N**	y	y	y	**N**	y	y	y	y
Minor	Vanishes for **afferent** system	y	y	y	y	**N**	**N**	**N**	y	**N**	**N**	**N**	y	y	**N**
	Vanishes for **efferent** system	y	y	y	y	**N**	**N**	**N**	**N**	y	**N**	**N**	**N**	**N**	**N**
	State-dependent	y	y	y	y	**N**	**N**	y	y	y	**N**	**N**	**N**	**N**	y
	Based on **symmetric** probability distance	**N**	**N**	**N**	y	**N**	**N**	n	**N**	**N**	**N**	**N**	**N**	**N**	**N**
	Intuitively **interpretatable**	2	2	2	2	2	0	2	2	2	0	1	0	0	0
	Computationally **tractable**	1	2	2	0	2	2	2	2	2	2	2	1	2	2

TABLE I: Properties of different integration measures. All but the third are desirable properties; capitalized N/Y (no/yes) indicate when an integration measure lacks a desirable property or has an undesirable one. The first four properties are generally agreed to be important, while the second set of four have been argued to be important by some authors. $\phi^{\mathrm{M}} \equiv \phi^{\mathrm{otuk}}$ and $\phi^{\mathrm{M}}_{kk'} \equiv \phi^{\mathrm{ofkk}}_{kk'}$.

Fig. 3. Table I from Tegmark 2016.

4. Artificial and Real Intelligence

4.1 AI Theory and Limitations

Turing machine is known to undecidable on some problems. A Turing machine is a "mathematical model of computation that defines an abstract machine,which manipulates symbols on a strip of tape according to a table of rules" (Minsky, 1967, as cited in 'Turing machine', n.d.). The turing machine can simulate any algorithm as long it can be constructed, "A Turing machine can do everything a computer can do" (Sipser, 2006, as cited in 'Turing machine', n.d.). The halting problem is to decide if the program with certain inputs on a Turing machine will run for infinite steps. Alan Turing has proved in 1936 that there is no universal algorithm to solve the halting problem, while the essence of this problem is self-reference, and the incompleteness of first order logic.

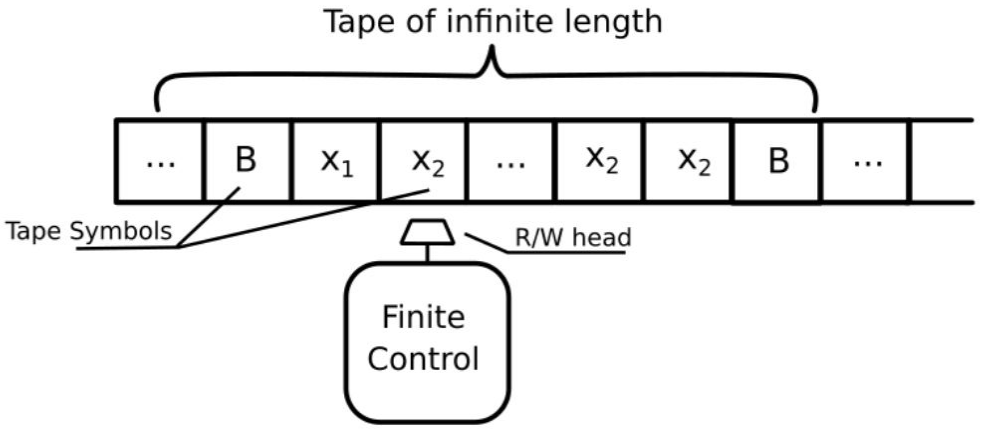

Fig. 4. An illustration of Turing machine. Retrieved from https://iq.opengenus.org/general-introduction-to-turing-machine/

This is also proved mathematically that a finite set of axioms cannot traverse all math relations. The Hilbert's second problem, as posed by David Hilbert in 1900, searches for a general proof if all propositions in a axiomatic system will be inconsistent. Further, he proposed that the consistency of a system can be proven in simpler ones and all mathematical consistency problems can be solved in basic arithmetic ('Hilbert's program',nd). However,

Gödel's incompleteness theorem (1931) has showed that even in a basic arithmetic system (with Peano axioms) it is impossible to deduct every true proposition and cannot thereby prove its own consistency, let alone other systems that includes basic arithmetics. In the computational function of intelligence with limited axioms, the incompleteness of algorithm (which contains Peano axioms) is a proof that it cannot find all the mathematical truth (all 'relations' or true propositions according to specific axioms) in the universe.

Therefore any Turing machine based AI system (modern computers), no matter how complicated, is confined to a limited space of creation (find new mathematical rules or patterns).

However, the world is physical other than merely algorithms. In physics, there are two types of description of the properties of objects: a certain (deterministic) physical unit or the probability (distribution) of some selected physical properties.

It has been argued for thousands of years whether the universe is deterministic or intrinsically random, the human (or 'real') intelligence are either: 1. Determined by the initial physical laws and specific environments in a deterministic universe and therefore is also confined to a finite set of information; 2. Having no limitation of finding new mathematical rules outside existing ones due to the intrinsic randomness of the physical laws.

There is a lack of definition of randomness in computational and mathematical form, here a definition of real random number generator is proposed (can be seen as the output from a device measuring an intrinsic random physical process): For time t and a number generator R(t), if there exist an algorithm S so that S(t) = R(t), then S is a 'full description' of R. The length (in digits) of algorithm S is denoted L(S). U(R) is the set of all 'full description' of R. If for any natural number N, and any L(S) that S belongs to U(R), we have L(S) > N, then R(t) is a real random number generator.

It is important to note that a real random generator is a reflection of physical intrinsic randomness in computational models, there is an inherent difference between real randomness and an 'enumeration' machine. For example, consider a book of 500 pages written in English, 'enumeration' machine will generate all the possible permutations and thus all that can be generated by a real random number (or letter) generator is equal to that of 'enumeration' machine. However, the algorithm of 'enumeration' machine is very simple for fixed amount of pages - basically a loop - and the program will have to change for different numbers of pages, while for the real random number generator, the L(S) is always infinite

and it can apply to whatever size of output space (and remains functional for reaching all possibilities).

This indicates that randomness by itself contains infinite information and mathematical axioms, which is crucial to the ability of (at least) innovation. Physical randomness, no matter which phenomenological layer gives such property, might be the very functional difference between computational system and real intelligence.

Even in the first deterministic assumption, AI systems should be designed to reflect more rules (information) of reality other than predetermined. The debate between symbolicism (to try represent all intelligent behavior and knowledge in symbolic form) and connectionism has almost come to an end with artificial neuron networks prevailing in numerous fields today. It is the fitting and integration of data that generate new and complex rules outside the original system that enhanced its intelligence. To put it another way, the real essence of training artificial neuron network is to locate a lower dimension manifold (which usually is the 'answer' to a question or a method to solve certain type of problems) in a higher dimension data space - the 'training' itself is a search with the parameters from outside (this is why the amount of data is important, since more data usually indicates higher precision).

4.2 Building a Functional 'General Intelligent' System

When it comes to the development of a 'general intelligence', one cannot exclude the reference from the only known example - the human brain. Although there is no definite evidence showing the exact physical mechanism of information processing in the brain to a readable level, it is clear that either the damp environment that consists of piles of neurons or the intrinsic nature of 'object reduction' that could happen on numerous microtubules in some patterns could provide sources of real randomness. A model based on real number generator and turing model of computation could be a reasonable attempt to establish functional 'real intelligence'. Here are two essential properties for such a system: (1) A general intelligent system should be able to give mathematical information (e.g. theorems) other than what has previously been included as axioms or anything derived from the axioms. (2) A general intelligent system should be no less than human in terms of the space of 'reachable' mathematical relations.

The first property is much easier to realize than the second one - which is not strictly described but necessary for most people to admit the 'generality' of the system. Simply put, the first step is to create a system (e.g. an algorithm) that make new mathematical rules with real random generators. In the real world organisms evolve in a hotbed of real randomness: the plasticity of molecules and cells (especially neurons) can easily 'extract' the randomness and extend beyond what has been coded in the system. Similarly, a genetic algorithm with artificial real random environment (sample set, input space,etc) should have the same capability. In fact, modern AI network (e.g. face recognition) is heavily relied on the big data - which is basically generated from the real random world. However, as the OR theory suggests, the randomness generating capability is perhaps more than 10^{16} operations per second for one single neuron, this might suggest that a general intelligence would need very powerful real random number generators. A framework that correlates outside input and intrinsic randomness will be necessary for GAI.

Here is a rough proof of why real randomness can give information outside a fixed system (or put it another way, solve any solvable problems within limited time): To simplify, we assume that any (mathematical) solvable question and its answer can be interpreted into a sequences of 0s and 1s without losing any information (once the answer is given, the correctness is automatically verified). Generally speaking, the length of the question and answer is unknown to the algorithm that is about to solve the problems. The process of solving this question is basically trying to give the answer's sequence. Now consider a deterministic algorithm, for any answer with length n, recall the 'numerator', one can easily solve the question with a loop. However, for any loop with n iterations, there exists a natural number M so that M>n, then this loop will not be able to find an answer with length of M. For an algorithm with real randomness, say a real random number generator giving out 0 or 1 values with 50% chances each, in the time period of t. For any answer of any question with the length of n, the chance of giving out the answer is 2^{-n} in a time length of nt, so the expectation of time length required to render the answer is n 2^n t, which is solvable. While it is not proved here whether a deterministic algorithm with limited length can solve all problems, the capability of real randomness is obvious. (To clarify, the real randomness in this assumption is of cardinality aleph 0. In terms of questions including 2^aleph 0, aleph 1 or

higher, it is not discussed here because such questions - if unable to describe in finite length - is outside of the concern of practical AI systems)

4.3 Revelations from Biological Evolution

4 billion years of evolution has shown how organisms accumulated changes and adapted to the environment. For example, human vision is well 'tuned' for perceiving the powerful part of the spectrum of sunlight, a tiny fraction of all spectrum allowed in the universe. Moreover, as sight is a obvious survival necessity, it is also adapted to the transmission of water, which has more constraints as the optimal spectrum is much narrower than that of the sun radiation.

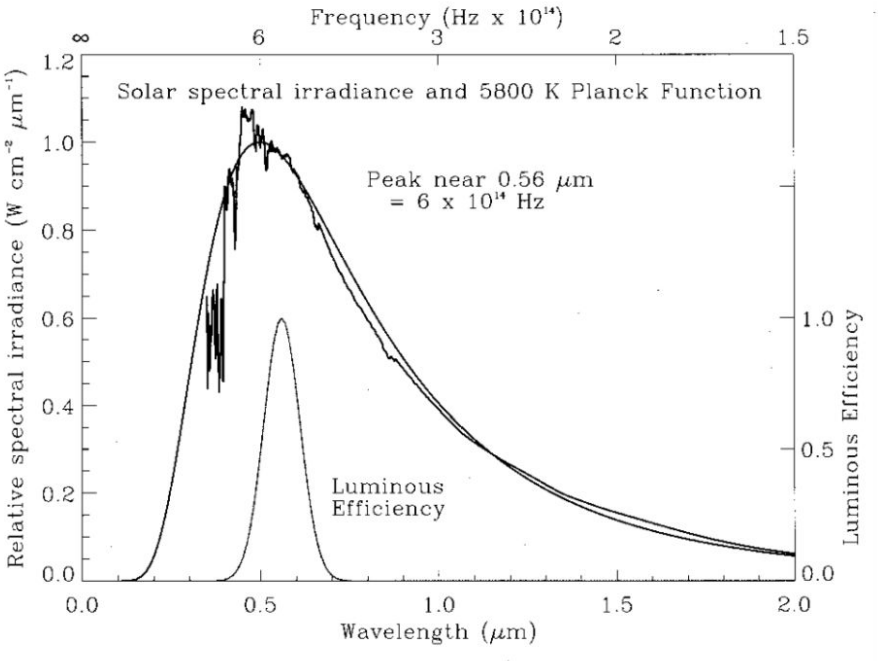

Fig. . The solar spectrum plotted in wavelength units peaks near 500 nm. Also shown is an approximate fit of a 5800 K Planck function that has been scaled to match the solar spectrum. This shows that the solar spectrum is roughly Planckian in the optical part of the spectrum. The luminous efficiency of the eye peaks at 560 nm. All three curves appear to peak near 500–560 nm, a wavelength region generally perceived as being green.(Soffer, 1999)

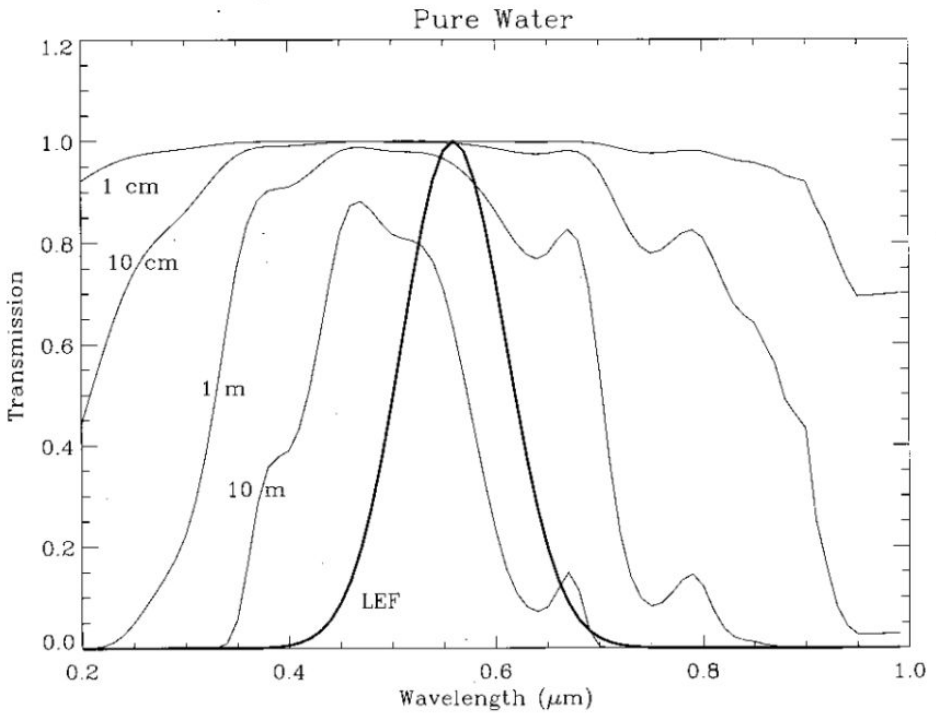

Fig.. The transmission of pure water compared with the luminous efficiency function of the eye. This calculation was done for absorption, and scattering was ignored. (Soffer, 1999)

The evolution of eyes itself is a intriguing subject as it is often provided as the (best) evidence of how 'impossible' evolution is, since the eye is too delicate to be 'evolved'. However, Nilsson and Pelger (1994) calculated "a pessimistic estimate of the time required for an eye to evolve", they "outlined a plausible sequence of alteration leading from a light-sensitive spot to a fully developed lens eye" and "the model sequence is made such that every part of it results in an increase of the spatial information the eye can detect". Then the number of generations required is calculated using " the amount of morphological change required for the whole sequence" (Nilsson, 1994). The approach is 'pessimistic', but the results showed that the evolution of an eye require "only a few hundred thousand generations" (Nilsson, 1994). For example, if the length of a generation is one year (typical for fish), "it would take less than 364000 years for a camera eye to evolve from a light-sensitive patch" (Nilsson, 1994). Life indeed have the time and chance to evolve the eyes many times in a billion years.

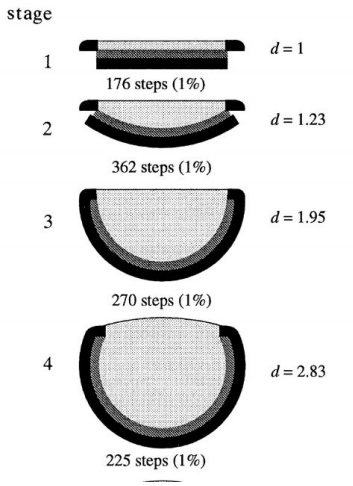

Fig. . Representative stages of a model sequence of eye evolution.(Nilsson, 1994)

Similar to the functionalities of life, the evolution of consciousness may also be a result of natural selection in terms of subjective experience (feelings tuned for function) as well as its corresponding automata mechanism (function tuned for survival). This would suggest that there are way more states of consciousness allowed than perceived in reality, as the evolution selects the very needed part from a unknown, immense space of consciousness.

5. Hypothesis of Background Consciousness Field

The efforts of trying to bring in qualia properties into an existing physical system have to assume that quale is intrinsic in existing forms of matter. But the space of qualia (whether it is reducible to the proto-quale or consists of numerous independent elements) is also unknown and may far exceed what human can perceive. As a result, such theory would either: 1. Claim numerous properties (including all elements of qualia) intrinsic in an existing (or known) physical process; 2. Claim that a single (or a few) quale is responsible for constructing all the subjective experience which also corresponds to a physical process (e.g. 'Objective Reduction'). However, the 'Hard Problem' required that the existence of qualia is as fundamental as matter in the deepest level, and all known physical laws needs to be revised for both of the assumptions.

In general, the qualia seem to be irreducible and inherently different in many cases for human observers (suggesting that there are many basic elements in reality) and the EM fields

(or other physical mechanism) do not seem to carry so many properties (huge redundancy for a physics theory), while natural selection tuned humans very well for several types of qualia for the 'human' automata, what could be the truth behind?

This project proposes the hypothesis of background consciousness field as a basic component of the reality, which consists of all possible qualia. The EM field (or perhaps through some process in objective reduction or other physical mechanism) interact with that field and this interaction will result in measurable consequences objectively (measured in physical method) and subjectively (the specification of a particular qualia from the background), that process will probably consume energy (corresponding to the level of valence). There is probably a principle of least action (like the one in physics) that dictates the max 'fitted' qualia to a physical system (which may account for the 'perfection' of subjective feeling to information processing). The conscious intelligence is different from a mechanical automata in that the input from the background consciousness field provide at least some additional information, if not a source of randomness (which lays the foundation of functional intelligence). The evolution of consciousness is a process by which organisms evolve toward more specific and higher arousal state of mind and toward a tuned integration and control for those states for some layer of the information processing in the brain. The EM field (or other physical mechanism) carries information rather than properties, and functions like a data cable of a monitor, the interaction creates frames of consciousness on the background 'screen'.

It is very important to search for the minimum physical requirement for consciousness, as a monumental achievement (to test this hypothesis) will be that one succeeds in creating an unprecedented subjective feeling in human brain by 'non biological' means. Building a phenomenological theory of how EM field or 'OR' selects (rather than 'create') qualia states would be the most practical approach at this stage. With a proven 'selection' theory we will then be ready for investigating the existence and properties of the consciousness field.

For the construction of a phenomenological theory, there are some hypotheses concerning which type/layer of physical mechanism could be candidates of a complete description of selecting qualia: (1) Classic EM theory. This is perhaps the oldest method of linking physics and consciousness, as the electric potential in the membrane is a key component of neuron activity. However, as stated in 'Orch OR' theory, the consciousness is not a static 'object' as we detect frames of consciousness related synchronized gamma waves. In addition, some

studies suggest that stimuli of magnetic field in particular section of the thalamus could trigger special experience (e.g. religious feeling of holiness). Since both the change of potential across the membrane (for the brain waves itself is potential oscillation) and direct magnetic stimuli can be represented by the presence of magnetic field, it is plausible to assume the magnetic magnitude distribution in space is a selector of qualia. Compared with other hypotheses, this one is easier for experiment design, as one can design delicate (both temporal and spatial) stimuli of magnetic field on cellular and system level. (2) The imaginary part of wave function. This hypothesis state that quale is created (in our case, selected) by the imaginary part of any wave function of an object. With these assumptions, all the quantum mechanism in an organism will yield proto-consciousness and only the evolved structure (e.g. microtubules) will generate sensible consciousness. The advantage of this theory is that it is attached to every physical process so that any quantum physical system that is able to form a mathematical structure for consciousness can actually form a real one, i.e. it is possible to build such devices (e.g. quantum computer) for further test. Unlike 'Orch OR', this hypothesis (although phenomenological) doesn't require a explanation of OR, which is another problematic problem. (3) Elementary particles. This hypothesis suggests that qualia are actually unknown elementary particles (or selected by some elementary particles in this case), since particles are essential an excited state in the field in quantum field theory, a finer investigation could be made to search for proper interactions in this level. The quantum vacuum field and the background consciousness field may be the same thing. An exchange of energy or transition, especially when photons and electrons are involved, could be a key in developing this theory. This is also a deeper version of the EM hypothesis. (4) Space-time structure. The 'Orch OR' theory had made this assumption but there was no information give as to how the proto-consciousness actually formed the different types of qualia. Instead, if we use a 'selection' approach, then it would probably be the tiny curvature of space-time and its detailed structure and dynamics that filter different qualia. Although some measures can be taken as to how gravitational field would affect this phenomenon - it is very hard to design an experiment on macro level as the gravitational field is usually too smooth and the delicate tiny difference required by OR is limited to the size of nano or even femto meters. A more practical approach is to find the correlation between patterns of beat frequencies and different qualia.

References

A general introduction to Turing Machine. (n.d.). Retrieved from https://iq.opengenus.org/general-introduction-to-turing-machine/

Brunden, K. R., Yao, Y., Potuzak, J. S., Ferrer, N. I., Ballatore, C., James, M. J., ... & Lee, V. M. Y. (2011). The characterization of microtubule-stabilizing drugs as possible therapeutic agents for Alzheimer's disease and related tauopathies. Pharmacological research, 63(4), 341-351.

Dermietzel, R. (1998). Gap junction wiring: anew'principle in cell-to-cell communication in the nervous system?. Brain Research Reviews, 26(2-3), 176-183.

Gauger E, Rieper E, Morton JJL, Benjamin SC, Vedral V. Sustained quantum coherence and entanglement in the avian compass http://arxiv.org/abs/0906.3725, 2011.

Hameroff, S. (1998). Anesthesia, consciousness and hydrophobic pockets—A unitary quantum hypothesis of anesthetic action. Toxicology letters, 100, 31-39.

Hameroff, S. (1998). Anesthesia, consciousness and hydrophobic pockets—A unitary quantum hypothesis of anesthetic action. Toxicology letters, 100, 31-39.

Hameroff, S. R. (2006). The Entwined Mysteries of Anesthesia and ConsciousnessIs There a Common Underlying Mechanism?. Anesthesiology: The Journal of the American Society of Anesthesiologists, 105(2), 400-412.

Hameroff, S., & Penrose, R. (2014). Consciousness in the universe: A review of the 'Orch OR'theory. Physics of life reviews, 11(1), 39-78.

Hilbert's program. (n.d.). Retrieved from https://en.wikipedia.org/wiki/Hilbert%27s_program

Hilbert's program. (n.d.). Retrieved from https://en.wikipedia.org/wiki/Hilbert%27s_second_problem

Johnson, M. E. Principia Qualia.

Nilsson, D. E., & Pelger, S. (1994). A pessimistic estimate of the time required for an eye to evolve. Proceedings of the Royal Society of London. Series B: Biological Sciences, 256(1345), 53-58.

Ouyang, M., & Awschalom, D. D. (2003). Coherent spin transfer between molecularly bridged quantum dots. Science, 301(5636), 1074-1078.

PENROSE, R. (2000). Wavefunction collapse as a real gravitational effect. In Mathematical physics 2000 (pp. 266-282).

Penrose, R. (2006). The road to reality: A complete guide to the laws of the universe. Random house.

Penrose, R., Gleiser, R. J., Kozameh, C. N., & Moreschi, O. M. (1993). Gravity and quantum mechanics. General relativity and gravitation, 13, 179-89.

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. Biosensors and bioelectronics, 47, 141-148.

Sahu, S., Ghosh, S., Hirata, K., Fujita, D., & Bandyopadhyay, A. (2013). Multi-level memory-switching properties of a single brain microtubule. Applied Physics Letters, 102(12), 123701.

Soffer, B. H., & Lynch, D. K. (1999). Some paradoxes, errors, and resolutions concerning the spectral optimization of human vision. American Journal of Physics, 67(11), 946-953.

Tegmark, M. (1996). A method for extracting maximum resolution power spectra from microwave sky maps. Monthly Notices of the Royal Astronomical Society, 280(1), 299-308.

Tegmark, M. (2015). Consciousness as a state of matter. Chaos, Solitons & Fractals, 76, 238-270.

Tononi, G., Boly, M., Massimini, M., & Koch, C. (2016). Integrated information theory: from consciousness to its physical substrate. Nature Reviews Neuroscience, 17(7), 450-461.

Turin, L. (1996). A spectroscopic mechanism for primary olfactory reception. Chemical senses, 21(6), 773-791.

Turing machine. (n.d.). Retrieved from https://en.wikipedia.org/wiki/Turing_machine

Tuszyński, J. A., Hameroff, S., Satarić, M. V., Trpisova, B., & Nip, M. L. A. (1995). Ferroelectric behavior in microtubule dipole lattices: implications for information processing, signaling and assembly/disassembly. Journal of Theoretical Biology, 174(4), 371-380.