



An Integration of Deep Learning and Neuroscience for Machine Consciousness

By Ali Mallakin

Abstract- Conscious processing is a useful aspect of brain function that can be used as a model to design artificial-intelligence devices. There are still certain computational features that our conscious brains possess, and which machines currently fail to perform those. This paper discusses the necessary elements needed to make the device conscious and suggests if those implemented, the resulting machine would likely to be considered conscious. Consciousness mainly presented as a computational tool that evolved to connect the modular organization of the brain. Specialized modules of the brain process information unconsciously and what we subjectively experience as consciousness is the global availability of data, which is made possible by a non modular global workspace. During conscious perception, the global neuronal work space at parieto-frontal part of the brain selectively amplifies relevant pieces of information. Supported by large neurons with long axons, which makes the long-distance connectivity possible, the selected portions of information stabilized and transmitted to all other brain modules. The brain areas that have structuring ability seem to match to a specific computational problem. The global workspace maintains this information in an active state for as long as it is needed. In this paper, a broad range of theories and specific problems have been discussed, which need to be solved to make the machine conscious. Later particular implications of these hypotheses for research approach in neuroscience and machine learning are debated.

GJCST-D Classification: 1.2.6



AN INTEGRATION OF DEEP LEARNING AND NEUROSCIENCE FOR MACHINE CONSCIOUSNESS

Strictly as per the compliance and regulations of:



RESEARCH | DIVERSITY | ETHICS

An Integration of Deep Learning and Neuroscience for Machine Consciousness

Ali Mallakin

Abstract- Conscious processing is a useful aspect of brain function that can be used as a model to design artificial-intelligence devices. There are still certain computational features that our conscious brains possess, and which machines currently fail to perform those. This paper discusses the necessary elements needed to make the device conscious and suggests if those implemented, the resulting machine would likely to be considered conscious. Consciousness mainly presented as a computational tool that evolved to connect the modular organization of the brain. Specialized modules of the brain process information unconsciously and what we subjectively experience as consciousness is the global availability of data, which is made possible by a non modular global workspace. During conscious perception, the global neuronal work space at parieto-frontal part of the brain selectively amplifies relevant pieces of information. Supported by large neurons with long axons, which makes the long-distance connectivity possible, the selected portions of information stabilized and transmitted to all other brain modules. The brain areas that have structuring ability seem to match to a specific computational problem. The global workspace maintains this information in an active state for as long as it is needed. In this paper, a broad range of theories and specific problems have been discussed, which need to be solved to make the machine conscious. Later particular implications of these hypotheses for research approach in neuroscience and machine learning are debated.

I. SCIENTIFIC VIEW AND DEFINITION OF CONSCIOUSNESS

One of the difficulties with the study of consciousness is the lack of a universally accepted definition. Another struggle is that empirical scientists must deal with questions, which has no quantitative answer, and responses can touch upon unscientific subjects. The very act of thinking demonstrates the reality of one's existence and consciousness, which viewed as an individual's awareness of its internal states as well as the events in the surrounding environment. Recent research on consciousness has focused on understanding the neuroscience behind our conscious experiences. Some investigators have used brain scanning technologies to seek out specific brain modules or neuronal networks that may link to different conscious events[1,2]. Two major theories of consciousness, which will be discussed further through this paper, consist of

integrated information and global workspace theories. The first attempts to look at consciousness by learning more about the physical processes that motivate our conscious experiences[3]. This theory attempts to create a measure of the integrated information that forms consciousness and reveals that the level of integration represents the quality of an organism's consciousness. It explains the nature and source of consciousness and claims that consciousness is identical to a particular type of information, which requires physical realization, not merely functional, and which can be measured mathematically according to the Φ metric. The second theory suggests that we have a memory bank from which the brain draws information to form the experience of conscious awareness[4]. While integrated information theory focuses more on identifying whether an organism is aware, the global workspace theory offers a broader approach to understanding how consciousness works. My motivation to have useful models in artificial consciousness is toward the design of communication technology that is convenient in everyday human activities.

There is a long way to go in our understanding of consciousness as investigators continue to explore the different bases of consciousness such as the physical, social, cultural and psychological influences that contribute to our conscious awareness. Consciousness defined as a state of awareness which features apprehensiveness, subjectivity, experience through sensory perceptions, the state of wakefulness, the sense of ego, and the control of the mind with knowledge of thought processes [5-7]. Other most essential attributes of consciousness are curiosity and creativity, which could be a factor that distinguishes humans from other earth beings[8,9]. Intelligence is also another fundamental aspect of brain requirement and consciousness, which some beliefs like consciousness it is a property of a physical mind, not a metaphysical phenomenon[10]; however, not a sufficient condition for creativity [11]. Apart from humans, other earth beings also show certain levels of creativity[12]. In defining consciousness, the problem emerges from attempts that try to explain it in purely physical terms. That is why psychology and quantum mechanics have been integrated to challenge the materialistic view of consciousness [13].

Recent advances in artificial intelligence (AI) have revived the possibility that machines would

Author: West Coast Biomedius, Vancouver, British Columbia, Canada.
e-mail: WCBIOMED@TELUS.NET

eventually mimic all the brain's abilities, including consciousness. Deep learning networks, inspired by neuroscience, have led to the creation of artificial neural networks (ANN) that can even occasionally surpass human capacity [14,15]. These types of the network without having biophysical properties of real neurons have several neurobiological features, including nonlinear input-output functions, layers with converging projections, and modifiable synaptic weights. Advances in computer field now allow such networks to operate on complex problems with capacity that previously was a privilege of a real brain. Computations implemented by current deep-learning systems that correspond mostly to non conscious processes in the human brain, although it often considered as the highpoint of the brain and something impossible to achieve by machines, in this paper is claimed to be otherwise. This paper reviews the previous studies and much of what we know about how brains generate consciousness, what are the relationships among conscious, subconscious and non conscious brains and how these findings can be used in advance artificial intelligence. This paper reviews specific existing models of consciousnesses and presents the approaches that make a machine more conscious. Besides, it discusses the role and relationship between quantum mechanics (QM) and remote viewing (RV). This relationship between quantum non locality and conscious or subconscious brain functions is essential, as quantum processes in biological phenomena are becoming more apparent. Further, it discusses non-local consciousness or remote viewing (RV) as an application of consciousness that permits a viewer to describe details regarding a target that is inaccessible to the ordinary senses.

II. CURRENT THEORIES FOR MACHINE CONSCIOUSNESS

There are various models of machine consciousness, each with certain strengths and limitations. Machine consciousness was classified earlier into four groups [16], and later to five categories based on recurring matters on the fundamental issues that are most fundamental to consciousness[17]. The most recent classification includes a global workspace, information integration, an internal self-model, higher-level representation, and attention mechanisms[17]. Although various models presented, the creation of an intelligent machine has not been much less successful due to the computational issues and the inability to explain the application of high-level cognitive algorithms in terms of neuro-computational processing[18]. Mechanistic theory of consciousness hypothesizes that how a brain with an attention schema may have subjective awareness[19]. In the attention schema theory, consciousness viewed beyond philosophy,

towards developing basic properties that can engineer into machines. It is considered an essential part of the brain responsible for data processing where conscious is an internal model of attention. Supporters of this theory claimed that the attention schema theory provides a possible answer to the puzzle of subjective experience by which the brain computes a simplified model of the process and the current state of attention which is the basis of subjective intelligence[20].

Consciousness made by the application of ANN may be better explained by exploring the designs that allow the human brain to generate consciousness, then transferring those understandings into computer algorithms. The aim is to study aspects of the cognitive neuroscience of consciousness that may be pertinent for learning machines. Consciousness given by brain capacity seeks to understand its relationship to our objective world. It states the relationship between a cognitive system and a specific object of thought, such as a mental representation. Conscious information becomes globally available to us, which gives us the ability to recollect, act upon it, and speak about it. Having the data in mind among the immense range of thoughts that can become conscious at a given time, only the ones that are universally available constitute the content of consciousness of an objective world. In another hand, consciousness scan be reflexive or spontaneous, in such a way that the cognitive system can monitor its processing and obtain information about itself in a self-referential manner. This self-examination ability of consciousness is commonly known as meta-cognition, which is having self-awareness and the ability to conceive and make use one's knowledge and abilities[21].Some suggested that global and reflexive consciousnesses may constitute independent dimensions of conscious computations. Both can overlap physical substrates, and in the brain, both depend on the prefrontal cortex[22]. Theoretically, the two may come apart as there can be global consciousness without reflexive consciousness, such as when reportable processing is not accompanied by accurate meta-cognition, or reflexive consciousness without global one, such as when a self-monitoring operation unfolds without being consciously reportable [23]. Sepertae computations can be performed before we consider their synergy. Furthermore, many estimates involve neither of these two conscious and therefore called unconscious. What we pursue is a satisfying scientific theory of consciousness that predicts under which conditions any particular physical system, whether it is a complex network of neurons or electronic circuit board has awareness experience. Moreover, we need to investigate what makes the quality of experiences different and if they have a specific function. Such a theory will allow us to conclude which systems will experience anything. Experience has shown that without an approach that brings testable

predictions, any speculation based on our intuition would not be reliable.

III. THE RELATIONSHIP AMONG CONSCIOUS, SUBCONSCIOUS AND UNCONSCIOUS MINDS

Sigmund Freud psychoanalytic theory presented in the early twentieth century has retain edits hold on the shaping of views regarding the doctrine of the human mind despite the new advancements. Neuroscience has identified the physical brain mechanisms underlying subjective processes. Following those breakthroughs, consciousness could then be seen as a computational property associated with a certain level of information processing. Three computational levels may be distinguished based on our conscious, subconscious or unconscious mind. It is estimated that just 10% of the cognitive function is made up of conscious thought, which can direct our focus and able us to envision what is not real. Our subconscious is the storage point of recent memories, which needed for quick recall and holds current information that we use every day. The unconscious mind is where our memories and past experiences reside and include memories that have been repressed and those that have been merely consciously forgotten. The unconscious brain continually communicates with the conscious mind via the subconscious, which provides us with the meaning of our interactions with the world. It breathes through feelings, emotions, thoughts, sensations, and dreams. At unconscious mind algorithms process symbols blindly and, apparently without any awareness. For instance, our visual system blindly and unconsciously processes the images. By this mean, many of the brain's unconscious computations are rational computations, and unexpectedly any machine that attempts towards objectivity would be submitted to similar human-like deceptions.

Above the unconscious processing level, two higher levels of information processing are defined that correspond to primary and secondary consciousness [24]. The first level known as conscious access, which is a selective attribution of a piece of information, selected for its relevance and became the focus of additional processing. Second, is conscious self-representation, in which the cognitive system shows one or several representations of its knowledge, for instance, it may know what it is currently focusing on, or it made an error. Consequently, the system not only commits its resources to a specific piece of information but also recognizes that it knows. The assumption is that this self-knowledge is represented in the same format as the knowledge of other individuals, which allows this information to be shared with others[25,26].

Those definitions made it entirely possible to study consciousness experimentally. The experimental

approach may proceed with identifying a minimal experimental paradigm that allows to contrast visible and invisible stimuli [27,28]. Then, carefully quantify the subject's introspection, which is recognizing what it known. Self-examination defines the conscious of subjective perception and must, therefore, be recorded alongside other objective measures of brain activity. The capacity to report a piece of information to oneself or other scan be the next phase of the study, in which this sense of consciousness that is called report ability can be distinguished from other concepts such as attention, watchfulness, or self-consciousness. The last can be the use of modern neuro-imaging tools to compare the behaviors and brain activity patterns evoked by reportable and unreportable stimuli, thus uncovering the signatures of consciousness.

IV. SPECIFIC SIGNATURES OF A BRAIN CONSCIOUSNESS

Further study of consciousness in the human brain reveals that although subliminal stimuli can induce considerable activity in many human minds, conscious perception is associated with a set of specific signatures. The first sign is amplification and access to the prefrontal cortex, in which a conscious image is amplified and gains access to higher levels of representation, particularly in prefrontal and parietal cortices. Next would be tracing the propagation of conscious and unconscious images that shows unconscious activity can be strong in the cortex, yet perish in a brief time within higher cortical areas. This conscious image is amplified in a non-linear manner, in an event known as global ignition.

Nevertheless, in a short period brain activity becomes more stable when the stimulus is conscious than when it is not [29]. This conscious ignition is accompanied by increased in bidirectional exchanges of information in the human brain. During a conscious episode, the cortex activates at greater distances, and correlations of brain signals demonstrate this. The brain can spontaneously generate its patterns of circulated activity with constant changes, even in the absence of stimuli[30]. This resting state activity can partially predict the content of consciousness. Finally, neurons in prefrontal and anterior temporal cortex fire to a specific concept and do so only when the corresponding word or image is presented consciously. This activity acted as a signature of conscious perception and known as the all-or-none law that applies to the firing of neurons[31].

The findings of brain-specific signatures and explanation of generic consciousness are compatible with the Global Neuronal Workspace (GNW) hypothesis, which tries to address the serious question of the function of consciousness[27,28,32]. Consciousness appears to be required for specific operations, and it is not just a mere epiphenomenon. Subliminal information

is temporary, but conscious knowledge is stabilized and available for long-term thinking. Consciousness converts the incoming data and reduces it to a form that can be reported or stored, while unconscious processes compute with an entire probability distribution, and consciousness samples from it through the subconscious mind. Consciousness is also involved in routing information to other processing phases, which allows us to perform random chains of operations [33].

V. MACHINE CONSCIOUSNESS SHOULD BE SELF-ORGANIZING

The phenomenon of consciousness may require a self-organizing system, similar to the brain's physical structure, which current machines lack this critical feature. Adaptive self-organizing mechanisms can be designed to be as sophisticated as the human brain; however, at the moment we require a mathematical theory of computation for these types of system. Perhaps it's true that only biological machines can be sufficiently creative and flexible. That may suggest start working on engineering new biological structures that are or can become conscious. Conscious as a useful aspect of brain function likely to be helpful to a design of artificial-intelligence devices. Certain computational features can be considered, which if they are implemented, the resulting machine would be likely to be considered conscious, or at least more aware than what is currently available. The first to be considered is a workspace for global availability, the act of selecting and making a piece of information accessible for processing and decision-making by the whole system. Global availability highlights or draws attention to information that remained unconscious until that moment. Even though the brain possesses specialized modules that operate non-consciously and are dedicated to specific tasks, it also maintains a global neuronal workspace, where accurate information is selected and shared across all modules. According to the GNW hypothesis, consciousness evolved to break this modularity. The GNW can extract relevant information from virtually any brain module, and make it available to the entire organism, in which the prefrontal cortex appears to act as a central information sharing device. The information presents in this global workspace at any given time is known as conscious. Machines may benefit from a similar architecture for flexible information sharing, capable of broadcasting to the entire system a potentially relevant piece of information. It would be interesting to pursue this idea in the context of present-day machine-learning algorithms, which can make the best use of the broadcasted information. Another feature for a conscious machine is its inherent abilities: a database that contains a list of its apps, the kind of knowledge they possess, what goals they can fulfill, how fast can operate, and how likely they should be corrected.

A conscious machine should know when it is wrong or when it is uncertain about something. In the human brain, this corresponds to meta-cognitive knowledge that has been linked to the prefrontal cortex. In several ways, a computer could be equipped with similar functionality. Primarily, it could be equipped with statistical programs that do not just give an answer, but also compute the probability that this answer is correct. Subsequently, it can have an error-detection system, similar to the brain's error-negativity, which continually compares the current activity with prior expectations and instinctively reacts if the current behavior is likely to be wrong. Finally, this error detection device could be attached to a corrective mechanism that allows the system to continually looks for alternative ways to get the correct answer.

Another point which may be unique to humans is the ability to represent self-knowledge and have a theory of mind, which allows us to model other accounts and to use this knowledge to maximize the usefulness of information that we can provide to them [34]. Additional attributes that should be considered include producing a machine with a consciousness that has qualities, instincts, and creativity. Also, states of consciousness that can change experiences such as emotions and expression scan be considered. Current machines often lack such significance and a device that could simulate its user's mind would undoubtedly provide more relevant information. Algorithms that handle such recursive representations of other accounts are currently being developed [35,36]. On the other hand, certain human behaviors such as creativity and a sense of freedom, which do not appear to come from logic or calculations, cannot be found in machines.

Another viewing platform on consciousness comes from quantum theory, which is one of the most in-depth theories of physics. According to the orthodox Copenhagen Interpretation, consciousness and the physical world are balancing aspects of the same reality. When a person observes some aspect of the physical world, that person's conscious interaction causes a noticeable change. Since this theory takes consciousness as a given and not derived from physics, the Copenhagen Interpretation may be considered as a phenomenon that exists by itself, while it requires brains to become real. This view was popular with the pioneers of quantum theory such as Niels Bohr, Werner Heisenberg, and Erwin Schrödinger. A well-known example of this is the paradox of Schrödinger's cat, in which a cat is placed in a situation that results in it being equally likely to survive or die – and the act of observation itself is what makes the outcome inevitable.

The opposing view reveals that consciousness emerges from biology, just as biology itself emerges from chemistry which, in turn, arises from physics. This less broad concept of consciousness agrees with the neuroscientists view that the processes of the mind are

identical to states and operations of the brain. It also coincides with a more recent interpretation of quantum theory motivated by an attempt to rid it of paradoxes, the Many Worlds Interpretation, in which observers are a part of the mathematics of physics.

VI. PARTICLE AND CONSCIOUSNESS ENTANGLEMENTS

Since the past few decades, quantum processes in the brain have been given attention to explain consciousness and its enigmatic features. As functional quantum processes in biology are becoming more and more apparent, recent evidence suggests that quantum non locality occurs in conscious and subconscious brain functions. Remote viewing (RV) is a mental ability that allows a viewer to describe or give details about a target that is inaccessible to normal senses [37,38]. For example, a viewer might be asked to specify a location in a different part of the world, which a reviewer has never visited; or a viewer might be asked to describe an event that happened long ago without being said anything about the target. Therefore, RV deals directly with target information processed by human consciousness. QM deal with particles such as electrons and photons, in which the concept of particle entanglement refers to local measurements on a particle that will instantaneously interact with an entangled particle, without considering the distance and how far apart the particles are [39]. QM and RV both involve entanglement, in which particle entanglement can be described in QM and consciousness entanglement in RV. Entanglement in non-quantum or classical version arises in situations where we have partial knowledge of the state of two systems. In QM entanglement, the quantum information is encoded in such a way that none of the two qubits (quantum bit: unit of quantum) carries any clear information on its own as all of the data is encoded in their joint properties. Such entanglement is one of the counter-intuitive features of QM and leads to its inconsistencies. Consciousness entanglement in the RV process refers to the local intentions by a viewer on target information that will instantaneously interact with entangled target information, no matter how far apart the viewer and target are. For the viewer, the entangled experience is the target. RV experimental result shows that nonlocal precognitive connections are, in fact, part of our reality. Our memory fits with the simple linear model of time as the only reality based on our accepted experience. While the non-linear QM model has not been experimentally verified in physics laboratories, RV precognition experimental data support the notion of backward through time motion. Therefore, non-local interactions may also involve general aspects of space-time described by the theory of relativity.

VII. NEUROSCIENCE AND CONSCIOUSNESS

Consciousness is qualitative and coherent, in which for each conscious experience some qualitative entities have that experience and each piece of this conscious experience fits precisely into a global picture of our world [40]. As conscious experience depends on brain activity; therefore, neuroscience has an extensive contribution in explaining consciousness. Except for neuroscience, no other area can provide so much information and advances concerning consciousness, which focuses on the neural correlates of consciousness. Considering that not all neural correlates are explanatory, so identifying the correlates is the first step in the neuroscience of consciousness. A thorough study of consciousness through neuroscience requires that we understand the relevant neural properties at the right level of analysis. Understanding of consciousness will unquestionably emerge from neuroscience but cracking its mysteries will need the knowledge and discoveries of many other scientific disciplines. To understand the connection between brain activity and consciousness, we need to focus on neurons, study computational and spiritual models, and conduct a theoretical analysis. Here we will focus on identifying such principles by considering the activity of neurons. The effort is to address how neural properties may determine when a state is conscious and how they may explain what the content is of that conscious. As a phenomenal consciousness is not reducible to either of these cases, there is a limitation to empirical explanations of phenomenal consciousness as empirical explanations are fundamentally either structural or functional.

Explanation of consciousness remains in the neural action that presents conscious contents and the unknown element that changes a state with that content from unconsciousness to consciousness. A variety of theories can describe generic consciousness that offers a need for the presence of higher-level neural properties necessary or adequate for the consciousness of a given state. Global Neuronal Workspace is one of these theories, which suggests access into the neural workspace is essential and sufficient for a state to be consciousness. In contrast, supporters of Recurrent Processing Theory believe that frequent processing in sensory areas is necessary and enough for perceptual consciousness, and therefore admission to the Workspace is not necessary. In Higher-Order Theories, seem the presence of a higher-order state connected to prefrontal areas of a brain is necessary and sufficient for a phenomenal experience. Subsequently, recurrent processing in sensory areas or access into the workspace is not necessary. At last Information Integration Theories consider a form of integration of information as essential and adequate parts for a state to be conscious.

Most neuroscientists instead of explaining the existence of consciousness in the biotic domain, try to explain generic consciousness by identifying neural properties that can turn consciousness on and off and by determining the neural representational basis of conscious contents. Modern neuroscience of consciousness attempts to explain consciousness by focusing on neural correlates of consciousness (NCCs) that is an essential initial step in understanding consciousness [41,42]. Talk of correlates is deep-rooted in the neuroscience of consciousness, so one must remember that the goal is to find the subset of neural correlates that are explanatory, in answering particular questions. Neuroscientific theories that contribute to explaining specific and generic consciousness further discussed here. We realize that there are limits to neural explanations of consciousness, precisely because of the explanatory gap [43]. The reference to neural correlates here means neural explanatory correlate of consciousness.

The theories discussed here provide higher level neural properties that are necessary or sufficient for generic consciousness of a given state. The first is global neuronal workspace theory that is tied to the access of the brain architecture and assumes perceptual, aide-mémoire, attentional, valuational and motoric functions of a cortical structure, which involves workspace neurons with long-range connections linking arrangements. The global workspace in neural terms defined as long-range workspace neurons within different systems that can constitute the workspace, but they should not necessarily identify within the workspace. A subset of workspace neurons becomes the workspace when they exemplify specific neural properties. What determines which workspace neurons constitute the workspace at a given time is the activity of those neurons given the subject's current state. The workspace then is not a rigid neural structure but a rapidly changing neural network, typically only a proper subset of all workspace neurons. This theory can also provide an account of phenomenal consciousness in addition to access consciousness. As imaging data can be obtained that reveal widespread activation when consciousness is present, it can predict widespread activation of a cortical workspace network as it correlates with phenomenal conscious experience [44]. As we track phenomenal consciousness by access in the introspective report, therefore, widespread activity during reports of conscious experience correlates with both access and phenomenal consciousness. However, this correlation cannot state whether the observed activity in a report is the basis of phenomenal consciousness or access consciousness.

Other theories such as Recurrent Processing Theory has a different explanation and ties perceptual consciousness to recurrent activity in sensory areas and processing independent of the workspace, with

emphasizes on properties of first-order neural representation as explaining consciousness. It claimed that repetitive processing is necessary and sufficient for consciousness [45-47]. This theory focuses mainly on the visual sensory modality, it argued to apply to all of the senses, and further, it distinguishes conscious and unconscious information. These distinctions were made along the lines of different types of processing that happen within the visual cortex and include the feed forward, locally recurrent, and globally recurrent types. Recurrent processing occurs where sensory systems are highly interconnected and involve feed forward and feedback connections.

The well-known Lamme's theory of recurrent processing identifies four stages of normal visual processing that consist of: I) In superficial feed forward processing stage visual signals are directed locally within the optical system; II) In deep feed forward processing step visual signals have traveled further forward and they can influence action; III) In superficial recurrent processing phase information has traveled back into former visual areas, leading to local, repetitive processing; IV) During widespread intermittent processing information activates extensive regions [45-47]. Based on this theory recurrent processing is necessary and adequate for consciousness. Therefore, the visual state is conscious once a specific intermittent processing state attains the suitable visual circuitry. Consequently, the global neuronal workspace embraces that recurrent processing at Stage 4 is necessary for consciousness while the recurrent processing theory holds that repetitive processing at Stage 3 is sufficient. Base on this difference, recurrent processing theory affirms phenomenal consciousness without access by the global neuronal workspace. Each approach has its answer, which is different from the other, one requiring access but then the other denying it. The organizational challenge in testing for the presence of phenomenal consciousness independently of access remains a hurdle for both theories.

Higher-Order Theory reveals that one is in a conscious state if and only if one relevantly represents oneself as being in such a state [48,49]. It can be described by a conscious visual state of seeing an object if and only if one appropriately represents oneself being in that visual state. This higher-order state by representing the first-order state that represents the world results in having the first order state that is conscious. The justification for such theories is that if one is in a visual state but not aware of that state, then the visual state would not be conscious. Consequently, to be in a conscious state, one must be aware of it and represent it [50]. Higher-order theories tie high-order representations with activity in prefrontal cortex, which is assumed to be the area with the required higher-order representations. On certain higher-order theories, one can be in a conscious visual state even without any

visual system activity, as long as one represents oneself as being in that state.

Empirical tests of the higher-order theory against other accounts show the need to study the prefrontal cortex of a brain [51]. Based on the higher-order theory, lesions to prefrontal cortex should affect consciousness; therefore, testing the prefrontal cortex for consciousness is necessary [52]. Against higher-order theories, recent studies claim that patients with surgically removed prefrontal cortex maintain preserved perceptual consciousness [53], which gives support to recurrent processing theories, therefore, deem that prefrontal cortical activity is not necessary for consciousness.

At last, Integrated Information Theory emphasizes on the concept of integrated information to explain generic consciousness, and it assumes consciousness needs a grouping of elements within a system that has physical cause-effect power upon one another[54-56].It postulates that any conscious system must possess three properties, which include: I) The existence of consciousness advocates a system of mechanisms with a specific cause-effect power. II) The organizational nature of consciousness suggests that its mechanical elements within a system must have the ability to combine and that those combinations have cause-effect power. III) Since consciousness is informative; it must specify or differentiate one experience from others.

This theory claims that consciousness is identical to a certain kind of information, in which its realization requires physical, not merely functional integration, which can be measured mathematically according to the Φ metric. In principle, it proposes the Φ metric to quantify consciousness and attempts to explain the nature and source of consciousness by incorporating that the system carries integrated information if the useful informational content of the whole is greater than the sum of the informational content of the parts. If there is no partitioning where the summed informational content of the parts equals the whole; then the system carries integrated information, and it has a positive value for Φ . Intuitively, the interaction of the components adds more to the system than the parts acting alone.

This theory attempts to bring a balance by preserving the cartesian intuitions that experience is immediate, direct and unified, but on the other hand, it takes neuroscientific accounts of the brain for an understanding of what the nature of a physical system must be to be conscious [57].The methodology of the most developers of this theory that are neuroscientists involves characterizing the primarily subjective nature of consciousness and postulating the physical traits necessary for a system to realize it. The theory claims that these predictions agree with observations of the brain's physical realization of consciousness; therefore,

once the brain does not present the necessary attributes, it does not create consciousness. Encouraged by these predictions, it generalizes its claims beyond human consciousness to animal and artificial consciousness. Since it identifies the subjective experience of consciousness with objectively measurable dynamics of a system, it can be measurable in principle, and consequently, proposes the Φ metric to quantify consciousness [58]. Recent new studies present more novel and practical methods to estimate Φ from high-density EEG that can apply to different states of consciousness. The investigation of the EEG properties corresponding to Φ metric enables to gather information on large-scale network correlates of various states of consciousness [59,60].

VIII. CONCLUSION

This paper attempts to gather and present the features necessary to produce a machine consciousness with a more human-like experience for artificial systems. If the essential computational features that conscious brains possess and currently lack in machines can be implemented, then artificial systems that appear natural would be possible. In addition to the interpretation of consciousness as a metaphysical phenomenon, this paper presents its concrete definition based on the models and biological study of conscious brains. Here the author compared the various existing models of consciousness to highlight the essential and distinct features of what makes the machine aware. Even though, the definition of consciousness very much equals a biological perspective; the proposed machine consciousness is entirely computational and includes various phenomenological characteristics of consciousness. Based on the models and features presented in this paper, the organization of machine consciousness is a functional property and progressively approximated by incorporating features such as a workspace for global information sharing, up to date self-knowledge, statistical programs, error detection systems, and theory of mind, which the latter allows modeling other minds and use this information to maximize the usefulness of data. Also, machine consciousness needs to be an emerging phenomenon as the development of the human brain indicates that functional units such as prefrontal and anterior temporal cortex are responsible for the emergence of consciousness. As the central administration that controls and coordinates all processes, whether conscious or sub conscious is necessary for developing consciousness, the machine model of consciousness should mimic the biological systems functionally and retains a distinct architecture and basic algorithm for implementing consciousness. The conscious machine with its proposed features might be neither complete nor practically feasible at present; however, it should guide building models of conscious machines in the future.

We may not need to wait and see if computers become conscious as the design of more complicated computers has not yielded any satisfactory result. Instead, we may need to be more foresighted and proactively integrate machine consciousness in the new design processes to have immediate benefit from its function. Interesting philosophical and empirical studies can be done, which involve other fields in addition to computer sciences such as humanities, divinity, natural sciences, and psychology as an empirical science. It may be more useful to design a machine in a way that it concludes it has consciousness and can prove this conclusion. This intelligent machine then can use its self-model to regulate its data flow and understand the actions of others.

REFERENCES RÉFÉRENCES REFERENCIAS

1. Carhart-Harris, R.L., Leech, R., Hellyer, P.J., Shanahan, M., Feilding, A., Tagliazucchi, E., Chialvo, D.R., Nutt, D. (2014). The entropic brain: a theory of conscious states informed by neuro imaging research with psychedelic drugs. *Front Hum Neuro sci.* 3(8): 20.
2. Fins, J.J. (2008). Neuroethics and neuroimaging: Moving toward transparency. *American Journal of Bioethics.* 8:46–52.
3. Tononi G.I. (2005). Consciousness, information integration, and the brain. *Prog Brain Res.* 150: 109-26.
4. Baars, B.J., Franklin, S. (2009). Consciousness is computational: The LIDA model of global workspace theory. *International Journal of Machine Consciousness.* 1: 23-32.
5. Baars, B.J. (1997). In the theatre of consciousness. global workspace theory, a rigorous scientific theory of consciousness. *Journal of Consciousness Studies.* 4:292–309.
6. Revonsuo, A. (2006). *Inner presence: Consciousness as a biological phenomenon.* MIT Press.
7. Haladjian, H.H., Montemayor, C. (2016). Artificial consciousness and the consciousness-attention dissociation. *Consciousness and Cognition.* 45:210–225.
8. Hunter, J.A., Abraham, E.H., Hunter, A.G.: Goldberg LC, and Eastwood JD. (2016). Personality and boredom proneness in the prediction of creativity and curiosity. *Thinking Skills and Creativity.* 22: 48–57.
9. Lin, W-L., Shih, Y-L. (2016). The developmental trends of different creative potentials in relation to children's reasoning abilities: From a cognitive theoretical perspective. *Thinking Skills and Creativity.* 22:36–47.
10. Starzyk, J.A., Prasad, D.K. (2011). A computational model for machine consciousness. *Int. J. Mach. Conscious.* 3: 255-281.
11. Karwowski, M., Dul, J., Gralewski, J., Jauk, E., Jankowska, D.M., Gajda, A., Chruszczewski, M.H., Benedek, M. (2016). Is creativity without intelligence possible? a necessary condition analysis. *Intelligence.* 57:105–117.
12. Kaufman, A.B., Butt, A.E., Kaufman, J.C., Colbert-White, E.N. (2011). Towards a neurobiology of creativity in nonhuman animals. *Journal of Comparative Psychology.* 125(3): 255-272.
13. Brabant, O. (2016). More than meets the eye: Toward a post-materialist model of consciousness. *EXPLORE: The Journal of Science and Healing.* 12:347–354.
14. Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., Hubert, T., Baker, L., Lai, M., Bolton, A., Chen, Y., Lillicrap, T., Hui, F., Sifre, L., Van den Driessche, G., Graepel, T., Hassabis, D. (2017). Mastering the game of Go without human knowledge. *Nature.* 550:354-359.
15. Lake, B.M., Ullman, T.D. (2017). Tenenbaum JB, Gershman SJ. Building machines that learn and think like people. *Behav Brain Sci.* 40:e253.
16. Gamez, D. (2008). Progress in machine consciousness. *Consciousness and Cognition.* 17:887–910 (2008).
17. Reggia, J.A. (2013). The rise of machine consciousness: Studying consciousness with computational models. *Neural Networks.* 44: 112–131.
18. Reggia, J.A., Monner, D., Sylvester, J. (2014). The computational explanatory gap," *Journal of Consciousness Studies.* 21:153–178
19. Graziano, M.S.A., Webb, T.W. (2014). A mechanistic theory of consciousness. *International Journal of Machine Consciousness.* 6:163–176.
20. Graziano, M.S.A. (2017). *The Attention Schema Theory: A Foundation for Engineering Artificial Consciousness.* Front. Robot. AI.
21. Peters, F. (2013). Theories of Consciousness as Reflexivity. *The Philosophical Forum.* 44:341-372.
22. De Sousa, A. (2013). Towards an integrative theory of consciousness: Part 1 (neurobiological and cognitive models). *Mens Sana Monogr.* 11:100–150.
23. Peebles, G. (2013). Reflexive theories of consciousness and unconscious perception. *Philosophical Psychology.* 31:25-43.
24. Edelman, G. (1989). *The remembered present.* Basic Books: New York.
25. Bahrami, B., Olsen, K., Latham, P.E., Roepstorff, A., Rees, G., Frith, C.D. (2010). Optimally interacting minds. *Science.* 329: 1081-5.
26. Kerr, N.L., Tindale, R.S. (2004). Group performance and decision making. *Annu Rev Psychol.* 55:623-55.
27. Baars, B.A. (1989). *Cognitive theory of consciousness.* Cambridge, Mass.: Cambridge University Press.

28. Dehaene, S. (2014). *Consciousness and the Brain: Deciphering How the Brain Codes Our Thoughts* (Reprint edition). Penguin Books.
29. Schurger, A., Sarigiannidis, I., Naccache, L., Sitt, J.D., Dehaene, S. (2015). Cortical activity is more stable when sensory stimuli are consciously perceived. *Proceedings of the National Academy of Sciences of the United States of America*. 112, E2083-2092
30. Barttfeld, P., Uhrig, L., Sitt, J.D., Sigman, M., Jarraya, B., Dehaene, S. (2015). Signature of consciousness in the dynamics of resting-state brain activity. *Proceedings of the National Academy of Sciences of the United States of America*. 112:887-892
31. Quiroga, R.Q., Mukamel, R., Isham, E.A., Malach, R., Fried, I. (2008). Human single-neuron responses at the threshold of conscious recognition. *Proc Natl Acad Sci USA*. 105:3599-604
32. Dehaene, S., Changeux, J.P. (2011). Experimental and theoretical approaches to conscious processing. *Neuron*. 70:200-27
33. Sackur, J., Dehaene, S. (2009). The cognitive architecture for chaining of two mental operations. *Cognition*. 111:187-211
34. Sperber, D., Wilson, D. (1988). Communication and cognition. *Behavioral and Brain Sciences*. 10: 697-789
35. Baker, C.L., Saxe, R., Tenenbaum, J.B. (2009). Action understanding as inverse planning. *Cognition*. 113:329-49
36. Daunizeau, J., den Ouden, H.E., Pessiglione, M., Kiebel, S.J., Stephan, K.E., & Friston, K.J. (2010). Observing the observer (I): meta-bayesian models of learning and decision-making. *PLoS One*. 5(12):e15554
37. Brown, C. (2005). *Remote Viewing: The Science and Theory of Nonphysical Perception*. Farsight Press.
38. Marks, D. (2000). *The Psychology of the Psychic* (2nd Edition). Prometheus Books.
39. Einstein, A. Podolsky, B., and Rosen, N. (1935). Can Quantum-Mechanical Description of Physical Reality Be Considered Complete? *Physical Review*. 47: 777-780
40. Nagel, T. (1974). What is it like to be a bat? *Philosophical Review*. 83: 435-50
41. Crick F. and C. Koch. (1998). Consciousness and neuroscience. *Cerebral Cortex*. 8: 297-107
42. Crick F. and C. Koch. (2003). A framework for consciousness. *Nature Neuroscience*. 6: 119-126
43. Levine J. (1983). Materialism and qualia: The explanatory gap. *Pacific Philosophical Quarterly*. 64:354-61
44. Dehaene S. and Changeux JP. (2011). Experimental and theoretical approaches to conscious processing. *Neuron*. 70:200-27
45. Lamme, V.A.F, Roelfsema P.R. (2000). The distinct modes of vision offered by feed forward and recurrent processing. *Trends in neurosciences*. 23: 571-579
46. Lamme, V.A.F. (2006). Towards a True Neural Stance on Consciousness. *Trends in Cognitive Sciences*, 10: 494-501
47. Lamme, V.A.F. (2010). How Neuroscience Will Change Our View on Consciousness. *Cognitive Neuroscience*. 1: 204-220
48. Rosenthal, D.M. (1986). Two concepts of consciousness. *Philosophical Studies*. 49: 329-359
49. Rosenthal, D.M. (2005). *Consciousness and Mind*. Oxford: Oxford University Press
50. Rosenthal, D.M. (2002). Explaining Consciousness", In *Philosophy of Mind: Classical and Contemporary Readings*, David J. Chalmers (ed.), Oxford: Oxford University Press, 109-131
51. Hakwan, L. and Rosenthal, D.M. (2011). Empirical Support for Higher-Order Theories of Conscious Awareness. *Trends in Cognitive Sciences*. 15: 365-373
52. Kozuch, B. (2014). Prefrontal Lesion Evidence against Higher-Order Theories of Consciousness. *Philosophical Studies*. 167: 721-746
53. Boly, Melanie, Marcello Massimini, Naotsugu Tsuchiya, Bradley R. Postle, Christof Koch, and Giulio Tononi. (2017). Are the Neural Correlates of Consciousness in the Front or in the Back of the Cerebral Cortex? *Clinical and Neuro imaging Evidence*. *Journal of Neuroscience*, 37: 9603-9613
54. Tononi, G. (2004). An Information Integration Theory of Consciousness. *BMC*. 5: 42
55. Tononi, G. (2008). Consciousness as Integrated Information: A Provisional Manifesto. *The Biological Bulletin*. 215(3): 216-242
56. Tononi, G. (2015). *Integrated Information Theory*. *Scholarpedia*. 10 (1): 4164
57. Edelman, G. M., & Tononi, G. (2000). *A universe of consciousness: How matter becomes imagination*. New York, NY, US: Basic Books.
58. Tononi G., Boly M., Massimini M., Koch C. (2016). Integrated information theory: from consciousness to its physical substrate. *Nat. Rev. Neurosci*. 17: 450-461
59. Eagleman, Sarah L.; Drover, David R. (2018). Calculations of consciousness: electroencephalography analyses to determine anesthetic depth. *Current Opinion in Anesthesiology*. 31: 431-438
60. Hyoungkyu Kim., Anthony G. Hudetz, Joseph Lee, George A. Mashour, UnCheol L. and the ReCCognition Study Group. (2018). Estimating the Integrated Information Measure Phi from High-Density Electroencephalography during States of Consciousness in Humans. *Front Hum Neurosci*. 12: 42