

 <p>ISSN NO. 2320-5407</p>	<p>Journal Homepage: - www.journalijar.com</p> <h2 style="text-align: center;">INTERNATIONAL JOURNAL OF ADVANCED RESEARCH (IJAR)</h2> <p style="text-align: center;">Article DOI: 10.21474/IJAR01/4067 DOI URL: http://dx.doi.org/10.21474/IJAR01/4067</p>	 <p>INTERNATIONAL JOURNAL OF ADVANCED RESEARCH (IJAR) ISSN 2320-5407 Journal Homepage: http://www.journalijar.com Journal DOI: 10.21474/IJAR01</p>
---	--	--

RESEARCH ARTICLE

A CONCISE STUDY TOWARDS ENGINEERING CONSCIOUSNESS.

Mayukh Mukhopadhyay.

Communication, Media and Information ISU, Tata Consultancy Services Ltd., Kolkata-700156.

Manuscript Info

Manuscript History

Received: 24 February 2017
Final Accepted: 20 March 2017
Published: April 2017

Key words:-

IIT3.0, Phi, Intrinsic Information,
Shannon Information, GUIT

Abstract

Consciousness, as a phenomenon, has been a curiously neglected subject in the STEM community. In this review article, after briefly discussing prevalent philosophical challenges the author explores recent structured approaches towards quantitatively defining this neurobiological phenomenon in terms of stochastic models with a special emphasis on integrated intrinsic information. A new possibility of Grand Unified Information Theory (GUIT) is introduced as a future scope. The article also discusses various potential applications in medical science, autonomous vehicle and entertainment industry along with current technical roadblocks. The article concludes with the necessity of exploiting this phenomenon by incorporating it as a subject of mainstream STEM research.

Copy Right, IJAR, 2017,. All rights reserved.

Introduction:-

Consciousness has reached a common-sense consensual definition among various erudite schools of thoughts. It is defined as a phenomenon which begins in the morning when we wake up from a dreamless sleep, and it goes on all day until we fall asleep or die or otherwise become unconscious [1,4] [Fig 1]. Dreams are a form of consciousness on this definition [4]. Consciousness only vanishes during dreamless sleep or under general anaesthesia when, from our own intrinsic perspective, everything disappears and we experience nothing [4]. Consciousness consists of all the states of feeling or sentience or awareness [9].

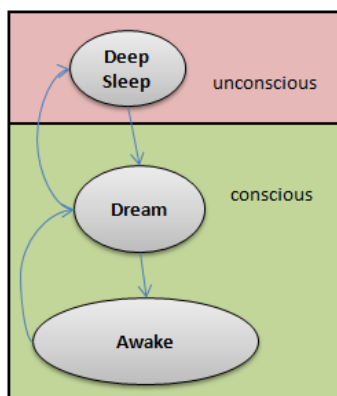


Fig 1:- Deep Sleep, Dream, Awakeness.

Corresponding Author:- Mayukh Mukhopadhyay.

Address:- Communication, Media and Information ISU, Tata Consultancy Services Ltd., Kolkata-700156.

All of our conscious states, without exception, are caused by lower-level neurobiological processes in the brain, and they are realized in the brain as higher-level or system features. It is analogous to the wetness of water [28].

INTER H₂O MOLECULAR INTERACTION : WETNESS :: INTER NEURAL INTERACTION : CONSCIOUSNESS

The wetness or liquidity is not an extra juice squirted out by the H₂O molecules [28]. It is a condition that the system is in. And just as the jar full of water can go from liquid to solid depending on the behaviour of the molecules, so our brain can go from a state of being conscious to a state of being unconscious, depending on the behaviour of the neurons [28]. A corollary can be drawn on the basis of above analogy that consciousness is an integrated and intrinsic form of mathematical pattern which arise due to a unique manner of interaction between specific groups of neurons.

Philosophical Challenges:-

Subjective-Objective Fallacy:-

If we look at the trend of using consciousness as a keyword in the English corpus for last five centuries, we see that it has been a relatively high-popularity subject compared to recently introduced concepts like artificial intelligence and neural networks [Fig 2].



Fig 2:- Keyword Trend in English Corpus from 1508 AD to 2008 AD

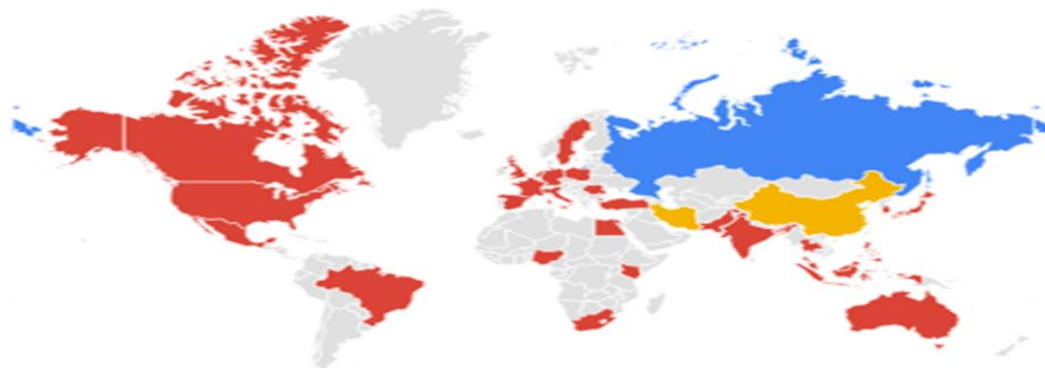
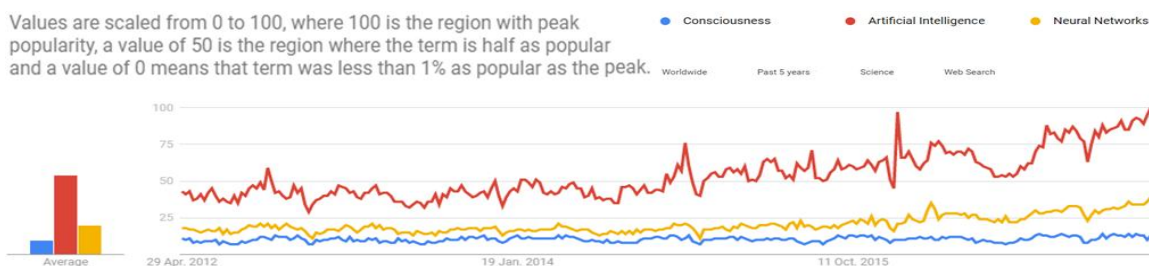


Fig 3:- Keyword Trend in Science Corpus from April 2013 to April 2017

However if we look at the recent trend in scientific literature [Fig 3], there is a popularity reversal between AI, neural networks and consciousness. This curious reluctance and hostility towards consciousness is a combination of two features of our intellectual culture that like to think they're opposing each other but in fact they share a common set of assumptions.

One feature is the tradition of religious dualism: Consciousness is not a part of the physical world. It is a part of the spiritual world. It belongs to the soul, and the soul is not a part of the physical world. That is the tradition of God, the soul and immortality. There is another tradition that thinks it is opposed to this but accepts the worst assumption. That tradition thinks that we are heavy-duty scientific materialists: Consciousness is not a part of the physical world. Either it does not exist at all, or it is something else, an illusion like sun rises in the east, but in any case it is not part of science. Most importantly the flawed argument goes that Science is objective, consciousness is subjective, and therefore there cannot be a science of consciousness.

Mind-Body Problem:-

The 17th century French philosopher René Descartes proposed the notion of "cogito ergo sum" ("I think, therefore I am"), the idea that the mere act of thinking about one's existence proves there is someone there to do the thinking.

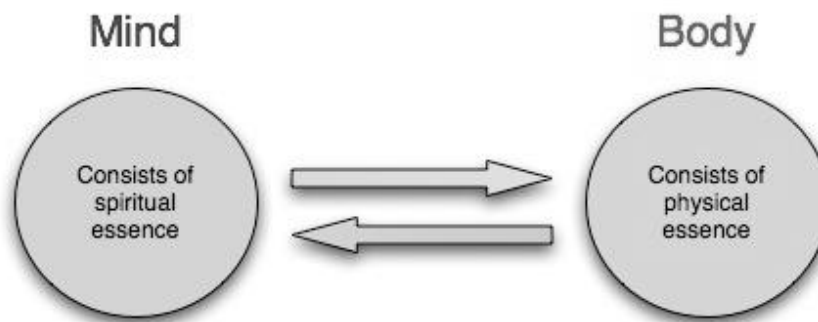


Fig 4:- Mind Body Problem

Descartes also believed the mind was separate from the material body. This is an age-old problem in philosophy known as the "mind-body problem"[Fig 4]. One quick way to state the problem is this: what are the relationship between the mind and the body; between the mental realm (the realm of thoughts, beliefs, pains, sensations, and emotions) and the physical realm (matter, atoms, neurons).

Panpsychism:-

Panpsychism is the claim that consciousness is not just a property of the brain, and not a property of some special spiritual kind of substance like the soul, but rather a property of everything in the universe.

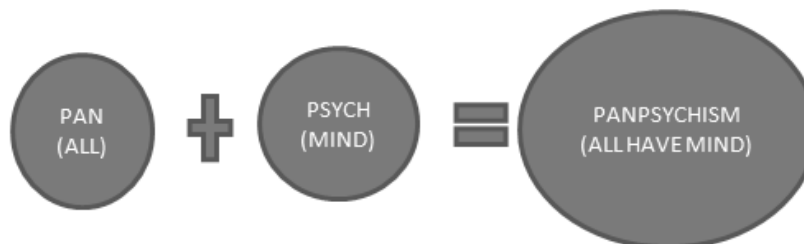


Fig 5:- Panpsychism

Here, consciousness is universal. Every system might have some degree of consciousness. This view states that every system is conscious, not just humans, dogs, mice, flies, but even microbes, a coffee mug and elementary particles. Even a photon has some degree of consciousness. The idea is not that photons are intelligent or thinking. But the thought is maybe photons might have some element of raw, subjective feeling, some primitive precursor to consciousness [1].

Revisiting Concept of Information:-

The word information derives from the Latin verb *informare*, which means 'to give form'.

Information as defined in information theory quantifies how accurately input signals can be decoded by knowing output signals transmitted across a channel, often noisy. Mutual information is the average statistical dependence between two sets of variables, such as the inputs and the outputs of a channel. The self-information contained in a single system state is simply the negative logarithm of its probability of occurrence. In general, this depends on the entropy of the inputs and on how well the inputs predict the outputs.

As clearly recognized by Shannon, the information of information theory is divorced from meaning: the sender and the receiver are assumed to know what the messages mean, and information theory is only concerned with how reliably and efficiently they can be transmitted across a channel. While Shannon's information can be evaluated between past states and current states of the same system, it is always assessed from the extrinsic perspective of an observer who assesses the statistical dependence between inputs and outputs. Therefore, Shannon's information could be called extrinsic information [20].

In the seminal paper "*Consciousness as Integrated Information: a Provisional Manifesto*", Giulio Tononi made a compelling argument that for an information processing system to be conscious, it needs to have two separate traits [15]:

1. *Information*: It has to have a large repertoire of accessible states, i.e., the ability to store a large amount of information.
2. *Integration*: This information must be integrated into a unified whole, i.e., it must be impossible to decompose the system into nearly independent parts, because otherwise these parts would subjectively feel like two separate conscious entities.

To characterize informativeness, Tononi used the Photodiode Thought Experiment. Imagine a photodiode responding only to light. The photodiode has two options, to remain off or to turn on. Each time there is light causally interacting with the photodiode in a certain way, it will turn on. Does the photodiode see light? According to Tononi, a crucial difference between us and the photodiode with respect to phenomenal experience is the space of possible discriminations it has. When a photodiode makes a discrimination between two alternatives, the detector in the photodiode generates $\log_2(2) = 1$ bit of information.

When we humans, or other animals, see 'light', by contrast, we discriminate between vast numbers of alternative things that we could have seen. In Tononi's words, we are implicitly being much more specific; we simultaneously specify that things are this way rather than that way (light as opposed to dark), that whatever we are discriminating is not colored (in any particular color), does not have a shape (any particular one), is visual as opposed to auditory or olfactory, sensory as opposed to thought-like, and so on. In contrast to the photodiode, we produce much more information when we see light than simply 1 bit of information. We exclude all the trillions of possible states that we have in our repertoire.

To understand why we also need integration, Tononi used the Camera Thought Experiment. Imagine a digital camera whose sensor chip is a collection of a million of binary photodiodes, each sporting a sensor and a detector [11,12,13].

With the camera, in contrast to the photodiode, it seems we get a huge amount of information. If we consider the camera as a single system with a repertoire of $2^{1,000,000}$ states then we have a million bits of information. When the camera takes a picture, it settles into one of $2^{1,000,000}$ states. If phenomenal experience was just a matter of discriminative information, then it seems that the camera should be very conscious. This is where the theory intuitively opposes. Tononi states that this camera chip is not an integrated entity since its 1 million photodiodes have no way to interact. The problem is that there is no special vantage point from which we might consider the camera as a single system or entity. What is needed is ultimately an intrinsic point of view associated with the camera chip as a whole. To understand such an objective measure we cut the camera sensor chip in two, such that it consists of two chunks of 500 thousand photodiodes each.

By doing this, we find that the information contained in the whole does not change at all. In fact, if we cut the camera sensor chip into 1 million pieces each holding its individual photodiode, the performance of the camera

would not change at all. The components in this system work independently of the whole. Hence the camera cannot integrate information while the individual photodiode does, even though it is of 1 bit.

Tononi's work has generated a flurry of activity in the neuroscience community, spanning the spectrum from theory to experiment, making it timely to investigate its implications for physics and mathematics [1].

In IIT the information content of an experience is specified by the form of the associated conceptual structure (the quality of the integrated information) and quantified by Φ_{\max} (the quantity of integrated information, phi-max). In IIT, information is causal and intrinsic: it is assessed from the intrinsic perspective of a system based on how its mechanisms and present state affect the probability of its own past and future states (cause–effect power). It is also compositional, in that different combinations of elements can simultaneously specify different probability distributions within the system. Moreover, it is qualitative, as it determines not only how much a system of mechanisms in a state constrains its past and future states, but also how it does so. Crucially, in IIT, information must be integrated. This means that if partitioning a system makes no difference to it, there is no system to begin with. Information in IIT is exclusive i.e. only the maxima of integrated information are considered. IIT assumes that for a system to be conscious, it must be a single, integrated entity with a large repertoire of highly differentiated states.

Since its inception, IIT has been further developed. In particular, IIT 3.0 considers both the past and the future of a mechanism in a particular state (it is so-called cause-effect repertoire) and replaces the Kullback-Leibler measure with a proper metric called EMD (Earth Mover Distance).

Information Theory	Integrated Information Theory
This theory was developed by CE Shannon.	This theory was developed by G Tononi
Here information is extrinsic by nature i.e. it is always assessed from the extrinsic perspective of an observer who assesses the statistical dependence between inputs and outputs.	Here Information is intrinsic by nature i.e. it is assessed from the intrinsic perspective of a system in terms of the differences that make a difference to it.
Here, information is divorced from meaning i.e. the sender and the receiver are assumed to know what the messages mean, and the theory is only concerned with how reliably and efficiently they can be transmitted across a channel.	Here information is meaning . It is causal, compositional, integrated and exclusive. More precisely, intrinsic integrated information is a maximally irreducible information structure (MICS) which is self-generated, self-referential, and holistic.
Here, information is a message transmitted across a channel.	Here, information is a shape in concept space.
Here, Kullback-Leibler divergence (KLD) measures the loss of information or reduction of uncertainty.	Here, Earth Mover Distance (EMD) is used as a true metric, that is sensitive to the relative distance among system states.
Here, the self-information contained in a single system state is simply the negative logarithm of its probability of occurrence.	Here, integrated information structure captures all the constraints over the past and future of a complex as determined by the state of its mechanisms.

Table 1:- Shannon's IT v/s IIT

Structured Approaches:-**Neurobiological Approach:-****Neural Correlates of Consciousness:-**

The NCC is defined as the minimum neuronal mechanisms jointly sufficient for any one specific conscious percept [18]. There are two possible interpretations of this definition, depending on whether we are referring to the specific content of consciousness or to the overall state of being conscious.

The *content-specific NCC* are the neurons (or, more generally, neuronal mechanisms), the activity of which determines a particular phenomenal distinction within an experience.

The *full NCC* are the neural substrates supporting conscious experiences in their entirety, irrespective of their specific contents. This is the union of the sets of content-specific NCC for all possible contents of experience.

Brain Regions:-

Tononi's functional MRI (fMRI) experiment revealed that the anatomical basis of the full NCC and content-specific NCC do not comprise the wide fronto-parietal network emphasized in past studies, but are primarily localized to a more restricted temporo-parietal-occipital hot zone with additional contributions from some anterior regions. At least some cortical areas in this posterior privileged zone, especially high-level sensory areas, are close to qualifying as content-specific NCC. Instead, the frontal cortex is involved in allocating attention and in task execution, monitoring and reporting.

In short, no single brain area seems to be necessary for being conscious, but a few areas, especially in the posterior cortical hot zone, are good candidates for both full and content-specific NCC. [4]

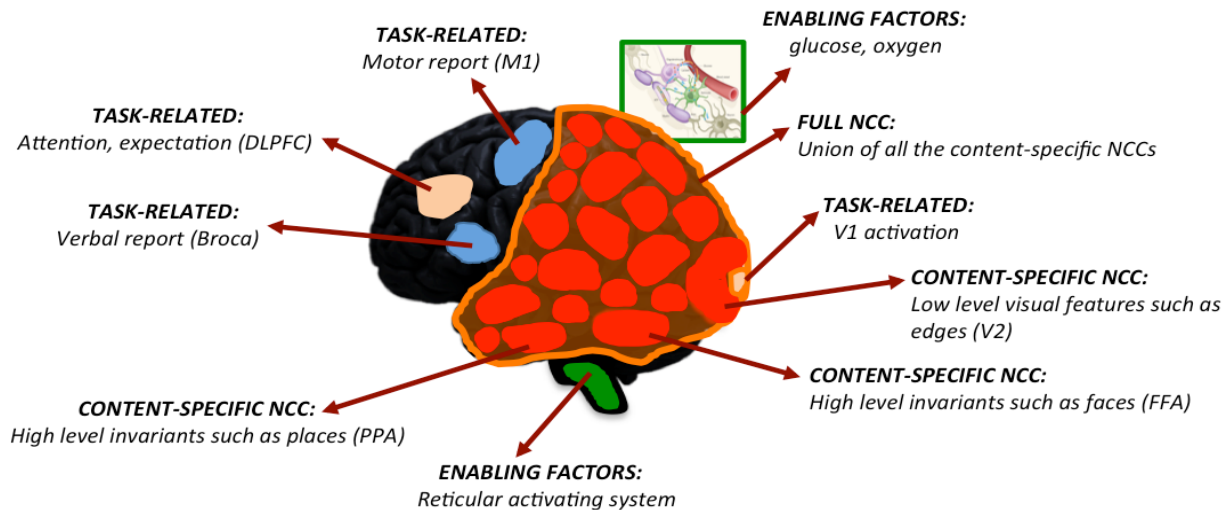


Fig 6:- Definition of neural correlates of consciousness (NCC) [6]

As depicted in the above figure, based on Tononi's review on evidences across lesion, stimulation and recording studies, Content-specific NCC (in red) directly contribute phenomenal distinctions (such as low level visual features, faces or places) to consciousness. The full NCC (in orange) is constituted by the union of all the content-specific NCC. Background conditions can be divided into physiological or neural processes that enable or modulate the activity of the full NCC and thus influence the level of consciousness (enabling factors, in green) - such as brainstem arousal systems, glucose or oxygen and task-related neural processes that modulate the activity of only some content-specific NCC (in beige) - such as attention, expectation, and possibly primary visual cortex (V1) activity. Other task related neural processes such as verbal or motor reports (in blue) often follow the experience itself. DLPFC: dorsolateral prefrontal cortex; V2: secondary visual cortex; FFA: fusiform face area; PPA: parahippocampal place area; M1: primary motor cortex [6].

Sleep Study:-

Although sleep study is apparently a digressed topic from the phenomenon of consciousness, it is a pervasive and universal behaviour. It occupies a third of our life, and is present in every animal species that has been studied. It is

also a fundamental behaviour. Even partial deprivation of sleep has serious consequences on cognition, mood, and health.

Tononi's hypothesis about the functions of sleep, the synaptic homeostasis hypothesis, claims that that sleep maintains synaptic homeostasis. In essence, sleep is the price we have to pay for plasticity, and its function would be the homeostatic regulation of the total synaptic weight impinging on neurons. The synaptic homeostasis hypothesis claims that plastic processes during wakefulness result in a net increase in synaptic strength in many brain circuits; especially during deep sleep, synaptic strength is globally downscaled to a baseline level that is energetically sustainable and beneficial for memory, performance and consciousness [15].

Physical Approach:-

Orchestrated OR Model:-

A provocative but seemingly plausible theory of consciousness was put forward by Penrose and Hameroff. They claimed that quantum coherence and the phenomenon of quantum wave function self-collapse, called objective reduction (OR), are essential to explain consciousness, and that these occur in cytoskeletal microtubules and other structures in neuron. Microtubules are particularly suitable for quantum effects due to their crystal-like lattice structure, hollow inner core, and organization of cell function.

They proposed that conformational states of tubulins are coupled to internal quantum events, and cooperatively interact with other tubulins. Their primary postulate is that macroscopic coherent superposition of quantum coupled tubulin conformational states occur throughout the brain and provides the binding that is essential for consciousness.

The emergence of microtubule quantum coherence corresponds to pre-conscious processing which grows (for upto 500 milliseconds) until the mass-energy difference among the separated states of tubulins reaches a threshold. At this point, the system must reduce and collapse to a single state that results in consciousness experience [2,27].

There were a few anomalous assumptions in the above theory, one of which implied that consciousness occurs throughout the entire brain. This has been corrected to the specific regions as discussed in the previous sections. Another fact is the brain operates at body temperature, it comprises of 60 per cent water, and electromagnetically, chemical and mechanically noisy which can severely shorten the time allowed for quantum computation (though microtubules may be the only candidate able to perform quantum computations at room temperature). Above all this theory was just a fancy postulate without any hard neurobiological evidence in support, as opposed to Tononi's work. Yet, this theory offered one of the early structured ways to approach towards understanding the esoteric phenomenon of consciousness.

Stochastic Quantification of PHI:-

Integrated information theory (IIT 3.0) starts from key properties of consciousness – the phenomenological axioms – and translates them into postulates that lay out how a system of mechanisms must be constructed to satisfy those axioms and thus generate consciousness.

IIT uses a mathematical stochastic model to quantify the amount of integrated information an entity possesses and thus its level of consciousness. The challenge is we cannot yet calculate the state of awareness for even the simple roundworm with current computers, let alone deal with the complexity of the human brain [5,29].

$$ei(x_1; P) = - \sum_{i=1}^k \sum_{\mu_0^{(i)}} p(\mu_0^{(i)} | x_1) \log p(\mu_0^{(i)} | \mu_1^{(i)}) - H(X_0 | x_1)$$

$$\Phi(x_1) = \min_P \frac{ei(x_1; P)}{v_P}$$

Fig 7:- Integration (PHI) measure in terms of Minimum Information Partition (MIP) [29]

Computational Approach:-**Global Neuronal Workspace (GNW) model:-**

Another promising theory suggests that consciousness works a bit like computer memory, which can call up and retain an experience even after it has passed.

Bernard Baars developed this theory, which is known as the global workspace theory. This idea is based on an old concept from artificial intelligence called the blackboard, a memory bank that different computer programs could access [5].

Anything from the appearance of a person's face to a memory of childhood can be loaded into the brain's blackboard, where it can be sent to other brain areas that will process it. According to Baars' theory, the act of broadcasting information around the brain from this memory bank is what represents consciousness.

IIT and global workspace theory are not mutually exclusive. The first tries to explain in practical terms whether something is conscious or not, while the latter seeks to explain how consciousness works more broadly.

Improved Measures:-

The study of Neural Correlates of Consciousness (NCCs) has become quite main-stream in the neuroscience community in recent years. To move beyond correlation to causation, neuroscientists have begun searching for a theory of consciousness that can predict what physical phenomena cause consciousness (defined as subjective experience) to occur. Dehaene reviewed a number of candidate theories currently under active discussion, including the Nonlinear Ignition model (NI), the Global Neuronal Workspace (GNW) model and Integrated Information Theory (IIT) [5]. Rapid progress in artificial intelligence is further fuelling interest in such theories and how they can be generalized to apply not only to biological systems, but also to engineered systems such as computers and robots and ultimately arbitrary arrangements of elementary particles.

The integration measure proposed by IIT is computationally infeasible to evaluate for large systems, growing super-exponentially with the system's information content. This has led to the development of various alternative integration measures that are simpler to compute or have other desirable properties. For example, Barrett & Seth [8] proposed an attractive integration measure that is easier to compute from time-series data, but whose interpretation is complicated by the fact that it can be negative in some cases. Casali et al. [24] used an integration measure inspired by complexity theory to successfully predict who was conscious in a sample including patients who were awake, in deep sleep, dreaming, sedated and with locked-in syndrome. Sitt et al. [25] suggest that state transition entropy correlates with consciousness. Griffith and Koch [26] have proposed defining integration of a system as the synergistic information that its parts have about the future, which appears promising although there does not yet exist a unique formula for it. Even Tononi et al. [21] has updated their integration measure twice through successive refinements of their theory. Despite these definitional and computational challenges, interest in measuring integration is growing, not only in neuroscience but also in other fields, ranging from physics and evolution to the study of collective intelligence in social networks [7].

Motivated by the growing interest in measuring integrated information in computational and cognitive systems, Max Tegmark [5] have presented a simple taxonomy of phi-measures where they are each characterized by their choice of factorization method (5 options), choice of probability distributions to compare (3×4 options) and choice of measure for comparing probability distributions (5 options). He classified all the integration measures revealed in this taxonomy by various desirable properties, as summarized in Table 3. When requiring the phi-measures to satisfy a minimum of attractive properties, the hundreds of options reduce to a mere handful, some of which turn out to be identical. All leading contenders are summarized in Table 2.

$\mathbf{A} = \mathbf{B}^T \mathbf{C}^{-1}$, $\mathbf{A}_b = \mathbf{B} \mathbf{C}^{-1} = \Sigma \mathbf{A}^T \Sigma_b^{-1}$, $\Sigma = \mathbf{C} - \mathbf{B}^T \mathbf{C}^{-1} \mathbf{B} = \mathbf{C} - \mathbf{A} \mathbf{C} \mathbf{A}^T$ and
 $\Sigma_b = \mathbf{C} - \mathbf{B} \mathbf{C}^{-1} \mathbf{B}^T = \mathbf{C} - \mathbf{A}_b \mathbf{C}_b \mathbf{A}_b^T = [\mathbf{C}^{-1} + \mathbf{A}^T \Sigma^{-1} \mathbf{A}]^{-1} = \mathbf{C} - \mathbf{C} \mathbf{A}^T \mathbf{C}^{-1} \mathbf{A} \mathbf{C}$.

\mathbf{C} is the data covariance matrix and \mathbf{B} is the cross-covariance between different times

Name	Definition	Formula for Gaussian variables
$\phi^{\text{otu}}(\phi^M)$	$I(\mathbf{x}^A, \mathbf{x}^B) - I(\mathbf{x}_0^A, \mathbf{x}_0^B)$	$\frac{1}{2} \log \frac{ \mathbf{T}_A \mathbf{T}_B \mathbf{C} }{ \mathbf{T}_A \mathbf{C}_A \mathbf{C}_B } = \frac{1}{2} \log \frac{ \Sigma_A \Sigma_B }{ \Sigma }$
ϕ^B	$I(\mathbf{x}_0^A, \mathbf{x}_0^B) - I(\mathbf{x}_0^A, \mathbf{x}_1^A) - I(\mathbf{x}_0^B, \mathbf{x}_1^B)$	$\frac{1}{2} \log \frac{ \mathbf{C}^2 \mathbf{T}_A \mathbf{T}_B }{ \mathbf{T}_A \mathbf{C}_A \mathbf{C}_B } = \frac{1}{2} \log \frac{ \mathbf{C} \Sigma_A \Sigma_B }{ \Sigma \mathbf{C}_A \mathbf{C}_B }$
$\phi^{\text{otum}}(\phi^{MD})$	$d_{MD}(\rho, q)$, $q_{ij} = \frac{p_{ij} p_{ij} p_{ij}}{p_{i-} p_{j-} p_{ij}}$	$\frac{1}{2} \left[\ln \frac{ \mathbf{C} }{ \Sigma } + \text{tr}(\hat{\mathbf{C}}^{-1} \mathbf{C} - \hat{\Sigma}^{-1} \Sigma - \mathbf{A}^T \hat{\Sigma}^{-1} \mathbf{A} \mathbf{C}) \right]$ for $\beta = 1$, $\hat{\mathbf{C}} \equiv \hat{\mathbf{A}} \mathbf{C} \hat{\mathbf{A}}^T + \hat{\Sigma}$
ϕ^{ots}	$I(\mathbf{x}^A, \mathbf{x}^B)$	$\frac{1}{2} \log \frac{ \mathbf{T}_A \mathbf{T}_B }{ \mathbf{T} }$
ϕ^{ofs}	$I(\mathbf{x}_1^A, \mathbf{x}_1^B)$	$\frac{1}{2} \log \frac{ \mathbf{C}_A \mathbf{C}_B }{ \mathbf{C} }$
ϕ^{ofu}	$-\sum_i p_{i-} \log \sum_j \frac{p_{ij} p_{ij} p_{ij}}{p_{i-} p_{j-} p_{ij}}$	$\frac{1}{2} \left[\ln \frac{ \mathbf{C}_A }{ \Sigma } + \text{tr}(\mathbf{C}_q^{-1} \mathbf{C} - n) \right]$, $\mathbf{C}_q \equiv \hat{\mathbf{A}} \mathbf{C} \hat{\mathbf{A}}^T + \hat{\Sigma}$
$\phi_{\text{sk}}^{\text{otk}}(\phi_{\text{sk}}^M)$	$\sum_i \frac{p_{i-}}{p_{i-}} \log \frac{p_{i-} p_{i-} p_{i-}}{p_{i-} p_{i-} p_{i-}}$	$\frac{1}{2} \left[\mathbf{x}_0^T (\hat{\mathbf{A}} - \mathbf{A}) \hat{\Sigma}^{-1} (\hat{\mathbf{A}} - \mathbf{A}) \mathbf{x} + \ln \frac{ \Sigma_A \Sigma_B }{ \Sigma } + \text{tr} \hat{\Sigma}^{-1} \Sigma - n \right]$
$\phi_{\text{sk}}^{\text{nak}}$	$\sum_i \frac{p_{i-}}{p_{i-}} \log \frac{p_{i-} p_{i-}}{p_{i-} p_{i-}}$	$\frac{1}{2} \left[\Delta \mathbf{m}^T \hat{\Sigma}_A^{-1} \Delta \mathbf{m} + \ln \frac{ \Sigma_A }{ \Sigma } + \text{tr} \hat{\Sigma}_A^{-1} \Sigma_A - n_A \right]$, $\Delta \mathbf{m} \equiv (\hat{\mathbf{A}}_A - \hat{\mathbf{A}}_A) \mathbf{x}_0^A + \hat{\mathbf{A}}_B \mathbf{x}_0^B$
$\phi_{\text{sk}}^{\text{pk}}$	$\sum_i \frac{p_{i-}}{p_{i-}} \log \frac{p_{i-} p_{i-}}{p_{i-} p_{i-}}$	$\frac{1}{2} \left[\Delta \mathbf{m}^T \hat{\Sigma}_A^{-1} \Delta \mathbf{m} + \ln \frac{ \Sigma_A }{ \Sigma } + \text{tr} \hat{\Sigma}_A^{-1} \Sigma_A - n_A \right]$, $\Delta \mathbf{m} \equiv (\hat{\mathbf{A}}_A - \hat{\mathbf{A}}_A) \mathbf{x}_0^A + \hat{\mathbf{A}}_B \mathbf{x}_0^B$
$\phi_{\text{sk}}^{\text{otk}}$	$I(\mathbf{x}_1^A, \mathbf{x}_1^B \mathbf{x}_0 = k k')$	$\frac{1}{2} \log \frac{ \Sigma_A \Sigma_B }{ \Sigma }$
ϕ^{nas}	$-\sum_i p_{i-} \log \sum_j \frac{p_{ij} p_{i-}}{p_{i-} p_{j-} p_{ij}}$	∞
ϕ_k^{nak}	$-\sum_i \frac{p_{i-}}{p_{i-}} \log \sum_j \frac{p_{ij} p_{i-}}{p_{i-} p_{j-} p_{ij}}$	∞
ϕ^{tps}	$-\sum_i p_{i-} \log \sum_j \frac{p_{ij} p_{i-}}{p_{i-} p_{j-} p_{ij}}$	∞
$\phi_k^{\text{tpk}}(\phi_k^{2.0})$	$-\sum_i \frac{p_{i-}}{p_{i-}} \log \sum_j \frac{p_{ij} p_{i-}}{p_{i-} p_{j-} p_{ij}}$	∞
ϕ^{mas}	$-\sum_i p_{i-} \log \sum_j \frac{p_{ij} p_{i-} p_{i-}}{p_{i-} p_{j-} p_{ij}}$	$\frac{1}{2} \left[\ln \frac{ \mathbf{C}_A }{ \Sigma_A } + \text{tr}(\mathbf{C}_q^{-1} \mathbf{C}_A - n_A) \right]$, $\mathbf{C}_q \equiv \Sigma_A + \mathbf{A}_A \mathbf{C}_A \mathbf{A}_A^T + \mathbf{A}_{AB} \mathbf{C}_B \mathbf{A}_{AB}^T$
ϕ_k^{mak}	$-\sum_i \frac{p_{i-}}{p_{i-}} \log \sum_j \frac{p_{ij} p_{i-} p_{i-}}{p_{i-} p_{j-} p_{ij}}$	$\frac{1}{2} \left[\Delta \mathbf{m}^T \hat{\Sigma}_A^{-1} \Delta \mathbf{m} + \ln \frac{ \Sigma_A }{ \Sigma } + \text{tr} \hat{\Sigma}_A^{-1} \Sigma_A - n_A \right]$, $\Delta \mathbf{m} \equiv (\hat{\mathbf{A}}_A - \hat{\mathbf{A}}_A) \mathbf{x}_0^A$
ϕ^{mps}	$-\sum_i p_{i-} \log \sum_j \frac{p_{ij} p_{i-} p_{i-}}{p_{i-} p_{j-} p_{ij}}$	$\frac{1}{2} \left[\ln \frac{ \mathbf{C}_A }{ \Sigma_A } + \text{tr}(\mathbf{C}_q^{-1} \mathbf{C}_A - n_A) \right]$, $\mathbf{C}_q \equiv \Sigma_A + \hat{\mathbf{A}}_A \mathbf{C}_A \hat{\mathbf{A}}_A^T + \hat{\mathbf{A}}_{AB} \mathbf{C}_B \hat{\mathbf{A}}_{AB}^T$
ϕ_k^{mpk}	$-\sum_i \frac{p_{i-}}{p_{i-}} \log \sum_j \frac{p_{ij} p_{i-} p_{i-}}{p_{i-} p_{j-} p_{ij}}$	$\frac{1}{2} \left[\Delta \mathbf{m}^T \hat{\Sigma}_A^{-1} \Delta \mathbf{m} + \ln \frac{ \Sigma_A }{ \Sigma } + \text{tr} \hat{\Sigma}_A^{-1} \Sigma_A - n_A \right]$, $\Delta \mathbf{m} \equiv (\hat{\mathbf{A}}_A - \hat{\mathbf{A}}_A) \mathbf{x}_0^A$
$\phi^{2.5}$	$\min \{\phi^{\text{nak}}, \phi^{\text{tpk}}\}$	∞
$\phi^{2.5'}$	$\min \{\phi^{\text{mak}}, \phi^{\text{mpk}}\}$	
$\phi^{2.5''}$	$\min \{\phi^{\text{otk}}, \phi^{\text{otpk}}\}$	

Table 2:- Integration (PHI) for different measures [5]

		$\phi^{2.5}$	$\phi^{2.5'}$	$\phi^{2.5''}$	$\phi^{3.0}$	ϕ^M	ϕ^B	ϕ^{MD}	ϕ_{sk}^M	ϕ^{otk}	ϕ^{otpk}	ϕ^{ots}	ϕ^{ofu}	ϕ^{nas}	ϕ^{mas}	ϕ^{xfk}
Major	Always non-negative	y	y	y	y	y	N	y	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	N	y	y
	Vanishes for deterministic system (drawback)	n	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	y	N	y	y	y	y
Minor	Vanishes for afferent system	y	y	y	y	N	N	N	N	y	N	N	N	y	y	N
	Vanishes for efferent system	y	y	y	y	N	N	N	N	N	y	N	N	N	N	N
	State-dependent	y	y	y	y	N	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	N
	Intuitively interpretable	2	2	2	2	2	0	2	2	2	2	0	1	0	0	0
	Computationally tractable	1	2	2	0	2	2	2	2	2	2	2	2	1	2	2

Table 3:- Properties of different integration measures [5]

Unfortunately, these most general integration measures are unfeasible to evaluate in practice, with the computational cost growing doubly exponentially with b , the number of bits in the system: they involve a Markov matrix of size $n = 2^b$, and they also involve minimizing over approximately $N = 2^n = 2^{2^b}$ bipartitions. Generalizing the pioneering work of Barrett and Seth [8], Tegmark [5] derived formulas for the Gaussian case that are exponentially faster, involving manipulations of a matrix whose size grows as $2b$ rather than 2^b with the number of variables b . Moreover, he showed how the second exponential can also be avoided using an approximation using graph theory, thus reducing the computational cost from doubly exponential to merely polynomial in the system size b .

Axioms and Postulates:-

The initial theory of IIT has evolved through stages 2.0 and finally into 3.0. The main tenets of IIT can be presented as a set of phenomenological axioms, ontological postulates, and identities. While the terms “axioms” and “postulates” are often used interchangeably, IIT 3.0 the distinction takes on an even stronger meaning: axioms are self-evident truths about consciousness – the only truths that, with Descartes, cannot be doubted and do not need proof (experience exists, it is irreducible). Postulates instead are assumptions about the physical world and specifically about the physical substrates of consciousness (mechanisms must exist, be irreducible), which can be formalized and form the basis of the mathematical framework of IIT.

The central axioms, which are taken to be immediately evident, are as follows [21]:

- **Existence:** Consciousness exists – it is an undeniable aspect of reality. Paraphrasing Descartes, “I experience therefore I am”.
- **Composition:** Consciousness is compositional i.e. highly structured. Each experience consists of multiple aspects in various combinations. Within the same experience, one can see, for example, left and right, red and blue, a triangle and a square, a red triangle on the left, a blue square on the right, and so on.
- **Information:** Consciousness is informative. Each experience differs in its particular way from other possible experiences. Thus, an experience of pure darkness is what it is by differing, in its particular way, from an immense number of other possible experiences. A small subset of these possible experiences includes, for example, all the frames of all possible movies.
- **Integration:** Consciousness is integrated. Each experience is strongly irreducible to non-interdependent components. Thus, experiencing the word “PARA” written in the middle of a blank page is irreducible to an experience of the word “PA” at the right border of a half-page, plus an experience of the word “RA” on the left border of another half page – the experience is whole. Similarly, seeing a blue triangle is irreducible to seeing a triangle but no blue color, plus a blue patch but no triangle.
- **Exclusion:** Consciousness is exclusive. Each experience excludes all others and at any given time there is only one experience having its full content, rather than a superposition of multiple partial experiences; each experience has definite borders i.e. certain things can be experienced and others cannot; each experience has a particular spatial and temporal grain i.e. it flows at a particular speed, and it has a certain resolution such that some distinctions are possible and finer or coarser distinctions are not.

To parallel the phenomenological axioms, IIT posits a set of postulates [21]. These lists the properties physical systems must satisfy in order to generate experience.

- **EXISTENCE:** Mechanisms in a state exist. A system is a set of mechanisms.
- **COMPOSITION:** Elementary mechanisms can be combined into higher order ones.

The next three postulates, information, integration, and exclusion, apply both to individual mechanisms and to systems of mechanisms as depicted in Table 4.

MECHANISM	SYSTEM OF MECHANISMS
Information	
Only mechanisms that specify differences that make a difference within a system count	
Cause-effect information (cei): How a mechanism in a state specifies the probability of past and future states of a set of elements (cause-effect repertoires)	Conceptual information (CI): How a set of mechanisms specifies the probability of past and future states of the set (conceptual structure)
Integration	
Only information that is irreducible to independent components counts	
Integrated information (ϕ, “small phi”): How irreducible the cause-effect repertoire specified by a mechanism is compared to its minimum information partition (MIP)	Integrated conceptual information (Φ, “big phi”): How irreducible the conceptual structure specified by a set of mechanism is compared to its minimum information partition (MIP)
Exclusion	
Only maxima of integrated information count (over elements, space, time)	
Concept (ϕ^{Max}): A mechanism that specifies a maximally irreducible cause-effect repertoire (MICE or quale “sensu stricto”)	Complex (Φ^{Max}): A set of elements whose mechanisms specify a maximally irreducible conceptual structure (MICS or quale “sensu lato”)

Table 4:- Key concepts and measures of IIT [21]

Potential Applications:-

Application of integrated information, as a measure of consciousness, can be endless. The most impacted sector would be neuro-scientific community where we can devise non invasive equipments to detect and monitor consciousness in patients who are paralytic or in coma. Definition of “brain-dead” would be revisited in terms of intrinsic information parlance.

Another impacting sector in the near future would be transportation and logistics. With the advent of self-drive autonomous cars, integrated information would play a crucial safety parameter to quantify the degree of self-awareness in an autonomous machine. Deep Analytics will witness a paradigm shift from artificial intelligence to machine consciousness.

An over-ambitious but plausible industry can rise up by exploiting the neural correlates of consciousness to generate personalized lucid dreams. This could be a new form of entertainment industry which would blur the line between virtual reality (VR) and augmented reality (AR).

Future Scope and Roadblocks:-

Present roadblocks are purely technical and can be addressed as an engineering problem. For a markovian model of consciousness, the time complexity for calculating the transition probability matrix (TPM) is the main problem, which Tegmark [5] has addressed and catered with a direction for solving in polynomial manner using graph theory. A more practical question involves exploring ways of generalizing and further improving this graph-theory-based approximation for exponential speedup.

As a future scope, heuristic measures and experimental approaches inspired by IIT's theoretical framework will make it possible to test some of the predictions of the theory [21]. Deriving bounded approximations to the explicit formalism of IIT 3.0 is also crucial for establishing in more complex networks how some of the properties scale with system size and as a function of system architecture [5]. PyPHI, a python library for measuring integrated information in simple logic gate structures is available which can be used to simulate complex systems and mimic animats [3,19].

From an engineering perspective, a possible design for qualia-scope to detect MICS using non invasive technique or a phi-meter to detect the degree of consciousness can be a future possibility.

Even though the author agrees with the definite distinction between intrinsic and extrinsic information w.r.t the frame of reference, an exciting possibility exists for development of a Grand Unified Information Theory (GUIT) which could explain the correlates between intrinsic and extrinsic information, if at all it exists.

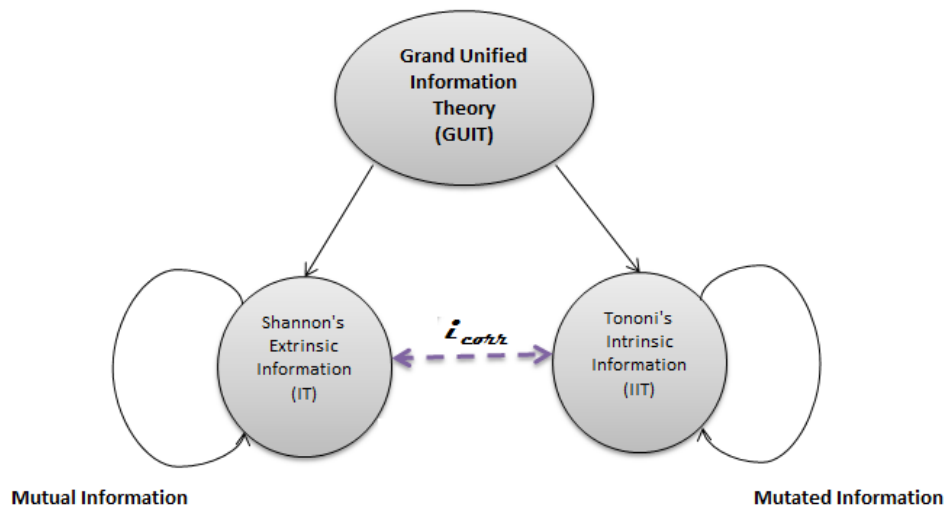


Fig 8:- Grand Unified Information Theory (GUIT)

Above figure [Fig 8] proposes a high level outline of this concept. Here we are introduced with two new terms, correlates of information and mutate information. Correlates of information (i_{corr}) represent the measure of

interaction between a message and a MICS, if at all this correlation exists. Mutate Information represents the degree of change introduced in a pre-existing MICS due to the interaction with the message. The term 'Mutate' is chosen deliberately as there seems to be some relation, again a future scope of study, between intrinsic information and evolutionary genetic mutation. To be specific, a hypothesis emerges, subject to proof, that consciousness in biological entity is an evolutionary step which got introduced to delay the inevitable process of mutation, which gradually occurs due to interaction with external competitive circumstances.

Conclusions:-

Standing today, we can safely conclude that consciousness has reduced down to an engineering problem. More specifically, if we consider integrated information, consciousness has become an optimization problem with polynomial complexity. Moreover, a veritable goldmine of data is becoming available in neuroscience and other fields, and it will be fascinating to measure integrated information for these emerging data sets [5]. It is interesting to consider whether one or several PHI-measures, proposed by Tegmark, can be rigorously derived from a small set of attractive axioms alone, in the same spirit as Claude Shannon derived his famous entropy formula [20]. Most importantly, can we extend the neural correlates of consciousness to the physical correlates of consciousness? In layman terms, can we make an autonomous machine conscious? For such a paradigm shift from machine intelligence to machine consciousness, it is a necessity to treat consciousness in an objective manner and introduce it as a main-stream STEM subject with STEM research grants and funding.

Acknowledgements:-

The author would like to dedicate this survey article to Late Prof. Satish Kumar, Ex Head, Department of Physics and Computer Science, DEI, Agra. Although the author never got the opportunity to meet him in person, the author respects his eruditeness and strong grasp towards neural networks [27] and belief in consciousness as a STEM (Science Technology Engineering Mathematics) topic. The author would like to thank John Rogers Searle and David John Chalmers for their inspiring lectures on TED talks on whose transcripts the introductory review notes has been based. The author would also like to thank G Tononi and Max Tegmark for showing the common masses a way to explore the enigmatic phenomenon of consciousness in an objective manner. The author would like to state that the survey done is sole responsibility of author and is not of the Organization to which the author is affiliated. Finally, the author is indebted to scrutinized proof editing, constant motivation and selfless support from his spouse, Mrityika. The author declares no conflicts of interest.

References:-

1. Max Tegmark, Consciousness as a State of Matter. Chaos, Solitons & Fractals March 17, 2015.
2. Stuart Hameroff, Quantum computation in brain microtubules? The Penrose-Hameroff 'Orch OR' model of consciousness, Phil. Trans. R. Soc. Lond. A (1998) 356, 1869-1896
3. Edlund JA, Chaumont N, Hintze A, Koch C, Tononi G, Adami C. Integrated information increases with fitness in the evolution of animats. PLoS Comput Biol. 2011; 7(10):e1002236. doi: 10.1371/journal.pcbi.1002236 PMID: 22028639
4. G. Tononi, Integrated information theory of consciousness: an updated account, Archives Italiennes de Biologie, 150: 290-326, 2012.
5. Tegmark M (2016) Improved Measures of Integrated Information. PLoS Comput Biol 12(11): e1005123. doi:10.1371/journal.pcbi.1005123
6. M Boly, G Tononi et. al, Are the neural correlates of consciousness in the front or in the back of the cerebral cortex? Clinical and neuroimaging evidence. (2017) doi.org/10.1101/118273.
7. Malone and Engel, Integrated Information as a Metric for Group Interaction: Analyzing Human and Computer Groups Using a Technique Developed to Measure Consciousness (2017), arXiv:1702.02462v1.
8. Barrett AB, Seth AK. Practical measures of integrated information for time-series data. PLoS computational biology. 2011; 7(1):e1001052. doi: 10.1371/journal.pcbi.1001052 PMID: 21283779
9. Tononi G, Edelman GM (1998) Consciousness and complexity. Science 282:1846-1851.
10. Seth AK (2009) Explanatory correlates of consciousness: Theoretical and computational challenges. Cogn Comput 1: 50-63.
11. Tononi G, Sporns O (2003) Measuring information integration. BMC Neurosci 4: 31.
12. Balduzzi D, Tononi G (2008) Integrated information in discrete dynamical systems: Motivation and theoretical framework. PLoS Comput Biol 4(6):e1000091.
13. Tononi G (2004) An information integration theory of consciousness. BMC Neurosci 5: 42.

14. Balduzzi D, Tononi G (2009) Qualia: The geometry of integrated information. *PLoS Comput Biol* 5(8): e1000462.
15. Tononi G (2008) Consciousness as integrated information: A provisional manifesto. *Biol Bull* 215(3): 216–242.
16. Tononi G (2005) Consciousness, information integration, and the brain. *Prog Brain Res* 150: 109–126.
17. Tononi G et. al (2016), Integrated information theory: from consciousness to its physical substrate, nature reviews, neuroscience.
18. Tononi G et. al (2016), Neural correlates of consciousness: progress and problems, nature reviews, neuroscience. MAY 2016. Volume 17, 307-321.
19. pyphi 0.8.1, Python library for computing integrated information. doi.org/10.5281/zenodo.12194
20. Tononi et. al (2014), Some differences between integrated information and Shannon information, *Text S3*, doi.org/10.1371/journal.pcbi.1003588.s004
21. Oizumi M, Albantakis L, Tononi G (2014) From the Phenomenology to the Mechanisms of Consciousness: Integrated Information Theory 3.0. *PLoS Comput Biol* 10(5): e1003588. doi:10.1371/journal.pcbi.1003588
22. Seth AK, Barrett AB, Barnett L. Causal density and integrated information as measures of conscious level. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*. 2011; 369(1952):3748±3767. doi: 10.1098/rsta.2011.0079 PMID: 21893526
23. Oizumi M, Amari Si, Yanagawa T, Fujii N, Tsuchiya N. Measuring integrated information from the decoding perspective. *PLoS Comput Biol*. 2016; 12(1):e1004654. doi: 10.1371/journal.pcbi.1004654 PMID: 26796119
24. Casali AG, Gosseries O, Rosanova M, Boly M, Sarasso S, Casali KR, et al. A theoretically based index of consciousness independent of sensory processing and behavior. *Science translational medicine*. 2013; 5(198):198ra105±198ra105. doi: 10.1126/scitranslmed.3006294 PMID: 23946194
25. Sitt JD, King JR, El Karoui I, Rohaut B, Faugeras F, Gramfort A, et al. Large scale screening of neural signatures of consciousness in patients in a vegetative or minimally conscious state. *Brain*. 2014; 137(8):2258±2270. doi: 10.1093/brain/awu141 PMID: 24919971
26. Griffith V, Koch C. Quantifying synergistic mutual information. In: *Guided Self-Organization: Inception*. Springer; 2014. p. 159±190. doi: 10.1007/978-3-642-53734-9_6
27. Satish Kumar (2013), *Neural Networks: A Classroom Approach*, 2e, McGraw Hill.
28. John Searle: Our Shared Condition – Consciousness, Jul 2013, TED Talk.
29. Christof Koch, A Theory of Consciousness Is complexity the secret to sentience, to a panpsychic view of consciousness?, *scientific american mind*, Jul-Aug 2009, 16-19.