Where is the Working Definition of Consciousness?

Contact: lkcl@lkcl.net. Date: 05 Mar 2025. Updated: 30 Mar 2025.
Ref: viXra:25??.???? v1

Abstract

The topic of Consciousness is astoundingly unique in that as a branch of Science there exists no clear definition of what Consciousness actually is. Tonini et al developed Integrated Information Theory and successfully synthesise the properties of consciousness to derive fascinating and ground-breaking insights, however stop short of deriving an actual working Definition. Beginning primarily from the Upanishads and Advaita Vedanta as the world's leading comprehensive and compact study on Consciousness, a first cut at a Definition of Consciousness is proposed and then refined with examples. Implications from the Definition: If "self" qualia are defined as the properties of thoughts, beliefs, memories, ideas, habis working paradigms etc. then the 'protection' mechanism is the self-referral loop, which bears a remarkable resemblance to an Error Checksum (ECC). Startling implications arise covering "Artificial" Intelligence, Computer Science, Epigenetics, Control Theory in Engineering, Vector Processing and Distributed Computing.

Contents

1 Working from a definition of Consciousness

Alan Turing created the Turing Test[27]:

> a test of a machine's ability to exhibit intelligent behaviour equivalent to that of a human... The results would not depend on the machine's ability to answer questions correctly, only on how closely its answers resembled those of a human.

An example of how selecting test subjects of both types could go awry is demonstrated by a human attempting to use ChatGPT to prove that the earth is flat [26]

Firstly: ChatGPT uses the wrong equation, and secondly the human lacks sufficient engineering knowledge to spot the mistake. Following Turing's Test, it would be hard to distinguish between the two, or even mistake the human as being "Artificial", as we (the Testers) could *mistakenly* reasonably assume Rationality on the part of both participants.

The bottom line is that the Analyst has to have a significantly high degree of Intelligence in order to correctly apply the Turing Test, and, also, have rather humorously-depressingly-realistic expectations of the selected human test subjects' Intelligence. Put clearer:

> In order to test for Intelligence one must first understand what Intelligence is.

And that involves having a *clear definition* of Intelligence. Maxwell's Demon neatly illustrates how simple observation may create a pressure differential under circumstances where none should naturally occur. Less-obvious but ironically illustrative is the case where the apocryphal millionth monkey eats the Works of Shakespeare immediately after typing it, not realizing the value of its randomly-typed output[24]

Likewise:

> In order to test for Consciousness one must first have a clear definition of what Consciousness is *and understand it.*

In Software Engineering it is deeply unwise to begin any large project without formally following strict steps. Simplified: those steps begin with an analysis phase, move to a specification document (defining the project and its scope), and only then to an implementation which, out of pure necessity, involves rigorous tests down even to the basic functions, all aimed at demonstrating or proving that the Software meets its clearly-defined specification. Thus, as all Software is explicit down to the most excruciatingly tedious level, then without clear and unambiguous definitions of every aspect of what is to be implemented it is literally impossible to deliver working software.

Yet astoundingly, exploration of Consciousness - and Conscious "AI" - is being carried out with no formal definition of what Consciousness actually is. Worse: the "A" in "AI" is a declaration by Western Science that Intelligence is somehow "Artificial". This mindset is now being transferred - ironically unconsciously - to "Conscious AI".

Correcting this mindset, and ensuring that vast funding resources are not wasted, would naturally require a definition of Consciousness. Dictionary definitions are hopelessly inadequate, and Internet searches refer to philosophical discussions, to "God" - anything but a clear and precise actual definition. Only by asking a different question, "What are the *properties* of Consciousness?" is there a hint of light, from which as a second step, with enough hints, can a first attempt be made at a workable definition.

Fascinatingly, it is Vedic literature - Advaita Vedanta and the Upanishads - that have the clearest discourse on Consciousness. Crucially, the discourse is in Sanskrit: a fully non-ambiguous language that can only result in miscommunication if either the writer *or the reader* misunderstands Sanskrit itself (dry diplomatic note: the level and quantity of misunderstanding, world-wide, is outside the scope of this document).

By contrast however, Phenomenology:

> seeks to objectively investigate the nature of subjective, conscious experience.

Which immediately contains an inherent catastrophic flaw: the assumption that subjective conscious experience is separate from the experiences: a subtle error that is made clearer by Phenomenology proponents' exploration of "Intentionality":

> "...intentionality, it being directed towards something, as it is an experience of or about some object."

The implicit assumption here is that Consciousness is a separate (objective) entity which can be "directed"[23].

Contrast this with the Upanishads all hinting that, using sentence-construction "subject-verb-object" to help illustrate:

- the "subject" (the self-aware being)
- observes (verb)
- *itself* (the object *is* itself),

in what becomes an endless infinite - recursive - loop:

- I observe I,
- observing I observing I,
- observing I observing I observing I,
- ...

2 The Properties of Consciousness

Synthesised from several sources, the properties are:

- Subjective awareness: knowledge of "self"
- The ability to interact with both an internal and external environment
- Knowing what it is like to be each of these three: (a) yourself (b) animate - i.e. knowing what it is like to interact with the environment (c) inanimate
- Being aware, sufficient to also be aware of that very same awareness: aware, and aware of the awareness.

Maharishi Mahesh Yogi in particular stated that Consciousness is "A state of pure being" - Pure Intelligence (Pure Creative Intelligence), Pure Knowledge, Pure Existence. Kashmiri Shaivism[25] also makes a clear statement that everything is part of one single large Consciousness, and hints again of awareness as key:

> There is only one Great Divinity, and it is the very inner Self of all creatures. It embodies itself as all things, full of unbroken awareness of three kinds: "I", "this", and "I am this."[69]

Further, from the Upanishads, an extensive collection, comes additional properties: the concept of being able to transition from "eyes of self" to "self", Advaita Vedanta as a study of

knowledge, and the identification of states of consciousness as wake, dream, sleep and transcendence.

Advaita Vedanta contains an extraordinarily detailed and comprehensive study of Epistemology, which if Consciousness is to be defined to the extent that it can be implemented in software, also needs to be incorporated. Advaita Vedanta identifies several types of knowledge, but it crucially also defines the structure as being of the form as:

- subject - a source
- cause or means - a link of some kind
- object - a destination

For a Software Engineer, the link between "subject" and "object" is of course a function. For anyone carrying out a study of language (grammar) the link between "subject" and "object" is of course known as a "verb".

For anyone studying Consciousness there are hints, particularly from the phrase "I am that", that self-observance - or self-awareness - must therefore imply that the "object" is also the "subject". This gives a huge hint as to one of the critical aspects of any definition of consciousness: feedback loops and interaction with the same are essential and key. From this same crucial aspect, it is quite clear that the majority of what is called "AI" catastrophically fails by design, as virtually all modern AI, being based on Directed Graphs, has no active integral feedback looping as part of the layering.

2.1 Knowledge types

There are different schools of Hindu philosophy, which advocate more or less "Pramanas" (types of knowledge). Nyana[22] only recognizes four Pramanas where Advaita Vedanta recognizes six.

Continuing with Advaita Vedanta Epistemology, the six types of knowledge are:

- The senses (which include "mind")
- Facts (or tenets)
- Inference: logical reasoning chaining: "A implies B and B implies C"
- Difference: in Boolean Logic and Software Engineering this is the inverse Exclusive OR Operation (XNOR)
- Analogy: A is to B as C is to D, which can be broken down to Set Theory Operations based on identified "attributes"
- Distinguishing between Cause and Effect: identification of a reflection not being the actual object itself. This is the "implies" operator A=>B, which in Boolean Logic is "B or NOT A".

Distinguishing cause and effect is illustrated by viewing an object in a mirror. Removal of the object results in the reflection disappearing, but removal of the *mirror* does not result in the *object* disappearing.

Remarkably a surprising number of the types of knowledge in Advaita Vedanta Epistemology map directly onto Boolean Algebraic operations. It is well-known for example that "Spot the Difference" image contests may be very rapidly solved by flashing up both images in quick succession, and the human eye is, from Evolutionary development, immediately drawn to the changes. Computer Software would perform an XNOR operation on every pixel of the digital images: any pixel that was the same would result in a ZERO and any pixel that was different a ONE. Searching for ONEs clearly locates the areas where the two images are different.

Also of note is hints of an ability to select from internal and external sources: "mind" being considered a sense, it is "mind" that provides stored "internal" information, as distinct from other "senses" which can definitely (albeit with some considerable processing layers in between) be considered "external".

2.2 Inference

The basis of Inference is the simple Boolean operation "A OR NOT B". It is part of a field called Propositional Logic[21]

Worth noting is that Inference Chains may be simplified through Exportation Rules[20]:

$$A => B => C \tag{1}$$

is equivalent to

$$A \text{ AND } B => C \tag{2}$$

Interestingly, the Nyana Wikipedia page on Inference is remarkably informative and detailed, so constitutes a valuable source. It outlines a *five* step process behind the Theory of Inference, citing the following example:

- There is fire on the hill (called Pratijñā, required to be proved)
- Because there is smoke there (called Hetu, reason)
- Wherever there is smoke, there is fire, e.g. in a kitchen (called Udāhārana, example of vyāpti)
- The hill has smoke that is pervaded by fire (called Upanaya, reaffirmation or application)
- Therefore, there is fire on the hill (called Nigamana, conclusion)

This does seem to be somewhat tedious, but to a Software Engineer it constitutes a clear and rigorous specification. Equally of value in the Nyana page is the warnings on common mistakes made by incorrect application of Inference Theory: the most common one being "If A=>B then B=>A" where it is well-known ad nauseam by Mathematicians that the Implies Operator is non-commutative.

"Where there is smoke there is fire" can be demonstrated as true, but "Where there is fire there is smoke" is demonstrably wrong, because smoke is generated by wet fuel: Methanol as a race fuel is known in NASCAR to be extremely dangerous as it burns (fire) entirely invisibly (zero visible sign of any kind except heat-haze, let alone any kind of smoke).

2.3 Analogy

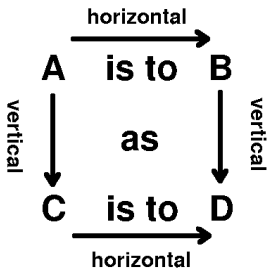

Figure 1: Horizontal and Vertical aspects of Analogy

Analogy deserves its own careful analysis. It involves Set Theory, as well as Boolean Logic. Imagine that each of A thru D are sets containing certain properties that are either True or False. Now imagine, as usual, that A and B are

laid out horizontally, and that C and D are directly underneath. Thus there are horizontal comparisons and vertical comparisons. Also it is further assumed that there is a unique "word" associated with each: a classic example being "A man B woman C king D queen". Properties in this example would be "human, male, female, royalty, commoner".

Comparing the set of properties of A to B (A XNOR B) will result in a subset of those properties which are the same. Likewise comparing C to D (horizontally), and also A-C and B-D (vertically). If the subset of properties when comparing A-B are identical to those of C-D, *and* those of comparing A-C are identical to those of B-D, then it becomes possible to *derive* (infer) the properties of one of the four items, as long as the properties of the other three are all known. Once deduction of the *properties* has been carried out, it is straightforward to perform a lookup of a suitable word or concept with those exact properties.

Interestingly, analogy is the focus of practical uses of Natural Language Processing, where large Sparse Matrices are constructed to define the relationship between words and associated properties. Each column becomes a Vector of properties, where the similarity of words may be mathematically calculated by calculating the Normalised dot product of the two vectors of properties (the cosine angle between the multi-dimensional vectors).[19]

The core algorithm takes four vectors, subtracts two pairs, and performs a search for the highest cosine angle[14]

```
# Compute cosine similarity between
# the vectors u and v
#u:(e_b - e_a)
#v:((w's vector representation) - e_c)
cosine_sim = find_cosine_similarity(e_b - e_a,
             word_to_vec[w] - e_c)
```

Each Dimension in each Vector above is the measure of any given specific "property". Thus it can be seen that by checking that the *difference* between two sets of properties are close, we have identified "similar" pairs: the definition of Analogy.

It should not take much further thought to see how a Neural Network, with the right design characteristics (such as a biological Spiking NN), could achieve the same type of calculation or an approximation of the same, although it is highly suspected that biological SNNs implement a different algorithm for Analogy which achieves the same end result. Interestingly, a Software Engineer or Mathematician should immediately recognise that both SNNs and computation of Dot Product are inherently parallel. Hong and Pavlik's approach confirms that something remarkably close to similarity (Dot product) can be achieved with what is termed Random Fourier Features[15]

Also worth noting in passing is that there are different categories of Analogy: positive analogy is described, above. Negative analogy is where the comparison of properties creates a set is found to be missing (X XNOR Y), but that the exact opposite of this comparison is the case for the other pair being compared (W XOR Z). Crucially it is noted that, in essence, Analogy builds on the use of "comparison" (one of the other types of knowledge in Advaita Vedanta).

Other refinements and subtleties occur when the properties being compared do not perfectly match. Another is that "plausibility" - confirmation or refutation of a "hypothesis" - requires an Analogy to be made between *internal and external knowledge*. Further reading is recommended in the Stanford Encyclopedia of Philosophy[13]

In conclusion on this brief analysis of Analogy, it would seem that Joseph Priestley was along the right lines to assert that analogy is a powerful tool in both discovery and philosophical investigation. However in and of itself, Analogy, as a type of *knowledge*, does not result in Consciousness.

3 Definition of Consciousness

The most important aspect of this proposed definition is that it is a hypothesis, and ironically, in a beautiful self-referral way, this paper *itself* notes that one of the tools utilised to confirm, reject or improve a hypothesis is: Analogy.

Thus, the definition must be compared against as many different expressions of Consciousness as is practical, synthesising aspects (properties) identified through Analogy.

Common attributes - properties - noted iteratively and corrected so far, are:

1. The ability to perform Boolean Logic (and from there optionally derive include or emulate basic Algebra - at least addition and subtraction)
2. The ability to perform differentiation and integration with respect to time.
3. The ability to read and write to and from internal storage capacity (memory)
4. The ability to select from any internal or external input (internal or external sources), at any given moment of time, for any combination of sources.
5. The ability to choose to write to external output, internal memory, or both or neither, or any permutation of the same, at any given moment of time.
6. Access to a source of true randomness, either as an additional input, or as an integral part of computation (whether Boolean, Algebraic, Differentiation or Integration), and a means of exerting control over the extent (effect) of that randomness.
7. Corrective regulatory real-time feedback looping that is integral to all decision-making throughout the entire system at every level (input selection, memory reading and writing, etc.).

This is a very specific and clear definition, which has remarkable similarities to the capabilities of any Computer which is running a truly pre-emptive multi-tasking Operating System (BSD, Linux) - a similarity that is explored below. Biological systems, which are known to operate at High-Order Critical Instability points (multiple self-corrective loops which are then themselves connected, usually and crucially in a non-Directed Graph), also meet the definition. Ant colonies and Bee colonies (as separate and distinct from the individual ants or bees themselves) also meet the definition.

Hankey[1] describes these as:

> The first important result derives from showing that critical instabilities contain an infinitesimal 'perfectly self-observing loop' of information

He goes on to point out:

> This occurs in the case of critical instability in flowing fluids, which occur at the critical Reynolds number [21,22]

Below we observe that a PID Controller, used to counter the effects of latency, surprisingly meets Hankey's definition, as the "loop" in PID Control involves differentiation and integration with respect to time of the input variable and the

output variable.

Remarkably, Anti-Virus software, by virtue of needing to check its own executable for virus infection, and also needs to monitor and stop real-time intrusion, including in what is supposed to be low-latency network traffic (web pages), also narrowly meets - or comes very close to meeting - the definition.

However before getting to these examples, it is important to go over how they were derived from the Properties of Consciousness.

3.1 Boolean Logic and Algebra

With the exception of Analogy, which depends on much more than Logic or Arithmetic operations, this was already partly covered above, from a simple analysis of Advaita Vedanta Epistemology (types of knowledge).

In some ways, adding Algebra to the definition is a bit of a cheat, because as any Computer Scientist or VLSI Engineer knows, Arithmetic operations may be built out of Boolean Logic Gates. Historically, the NOR Gate was used (CDC 6600), but when ASICs replaced transistor boards, the NAND gate was used. Either way: NAND or NOR is sufficient to build upon to create AND, OR and NOT. From there, XOR is constructed, and from all of those even the most complex Floating Point arithmetic units are ultimately constructed.

Within Biological Spiking Neural Networks it is well-known that both excitatory and inhibitory behaviour exists: combining these results in basic Boolean Logic operations, and as a result more complex logic - inherently parallel in form - occurs.

3.2 Differentiation and Integration wrt time.

In Biological SNNs, the ability to perform integration with respect to time is carried out at a fundamental basic level by every neuron. A chemical (neurotransmitter) accumulates inside the neuron, and its quantity increases with every synaptic firing. However, if there is an absence of firing, the neurotransmitter decays.

Additionally, neurons can either be inhibitory or excitatory (to other neurons), and the combination allows for Boolean Logic "Difference" between two states to be performed: the state that the neuron is in, and the state represented by its input. If at any given time the current state is compared with the previous state, this is pretty much the definition of Differentiation.

Combining these two aspects, then, we can say that every neuron has the crude ability to be part of performing both differentiation and integration with respect to time.

In Software Engineering and Mechanical Engineering, there is a concept known as PID Control: Proportional Integration Differentiation[28].

PID is essential for example in vehicle design to ensure that under unpredictable real-time driving, the engine is supplied with the right air-fuel mixture that will not cause catastrophic failure due to overheating. "Cruise Control" also uses PID to avoid the vehicle constantly oscillating around a target speed. Likewise in Gas boilers or Central Heating, PID Control helps with the significant latency involved in heating up of metal components. 3D Filament Printers also use PID Control to ensure that the heater block's temperature does not wildly fluctuate.

PID Control is *by definition* a form of real-time corrective feedback, that helps avoid oscillation and instability, correcting and accounting for delays between when a control signal is sent and when that control signal actually takes effect. Thus: wherever PID Control is used, even if implemented crudely, even in systems not normally considered as candidates for meeting the proposed definition of Consciousness, it should be considered significant.

So it is not just that a Conscious system must have Differentiation and Integration wrt: that must then be used to create real-time-adapting feedback that takes into account latency from input to output - aka "PID Control". Anyone who watches horses, chicks, or human babies attempting to stand for the very first time is witnessing PID Control at work with what is clearly the wrong initial settings. The fascinating bit is how quickly the horse, cow or chick adapts to the correct settings such that it can stable stand upright. Which begs the question: why do human babies take so much longer?[11]

3.3 Read and write storage

Even a cursory discussion of the need for internal storage (memory) and the ability to read and write to it should not be undervalued. It should be self-evident that without an ability to store information from the past, and without the ability to selectively recall that information, all of the other aspects of Advaita Vedanta Epistemology - Comparison, Inference, Analogy - all become flat-out impossible.

Note that the storage medium itself *and* the capability of actually reading and writing are needed, but without that storage medium itself - even if limited to storing only a single piece of information - Consciousness simply cannot arise.

3.4 Selection of input (source) and output (destination)

Selection of input (and whether to write to storage or output) becomes much more complex to discuss than the storage medium itself. In implementations of Computer Operating Systems, considerable effort is expended to ensure that unwanted data writing is prohibited, and that programs are likewise only allowed to read from permitted sources (files, memory, peripherals): Viruses and other Malware, just like their biological namesakes, are in a constant battle to bypass these critical restrictions.

3.5 Randomness

It turns out that randomness is an absolutely essential part of neural network information encoding. The technique is known as "whitening" and has the startling effect of increasing entropy. Pitti Weidmann and Quoy's approach shows that the biological equivalent of "Error Code Correction" from Computing corrects for randomness introduced by the crudeness of encoding[12].

The introduction of random noise uniformly increases the rate of firing across an entire Neural Network layer (or layers). Consequently, any rare events, such as the firing of a single cone from a single photon source, may be detected. Without the uniformly-distributed randomness, the rare event would be lost below the trigger-threshold of the neuron that the input was connected to.

3.6 Real-time regulation

Here is hints regarding the corrective feedback regulation being also involved. The regulation is necessary to keep the entire Neural Network "active": neither overloaded nor underloaded. A classic example of regulation would be the muscles in the eye controlling how much light reaches the retina. Thus when there is considerable light, causing the visual cortices to fire, the quantity of randomness may be reduced, and likewise during darkness, when there would otherwise be less activity, the quantity of randomness increased.

The effect then as noted by Pitti Weidmann and Quoy is that the network "adapts" to an approximation of logarithmic behaviour (interestingly using a biological equivalent of binary encoding), thus being able to cope with several orders of magnitude difference between dark and light.

3.7 Specific observation on PID Control

What exactly is PID Control?[28]

> a PID controller, which continuously calculates an error value $e(t)$ as the difference between a desired setpoint SP $= r(t)$ and a measured process variable PV $= y(t)$ such that $e(t) = r(t) - y(t)$, and applies a correction based on proportional, integral, and derivative terms.

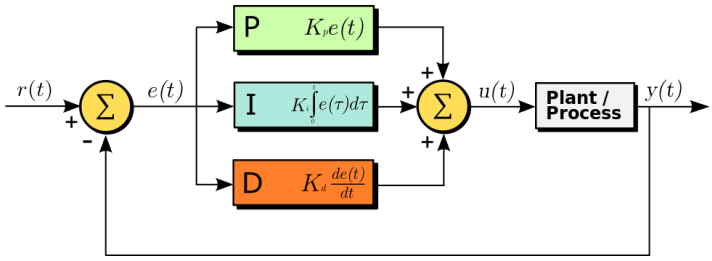

Figure 2: PID Control by Arturo Urquizo

Inherently and immediately it is clear that time is a critical input, and that it is the *difference* between the input and the measured output that drives the Control. To illustrate a key aspect of a PID Controller: setting P=1.0/I=0/D=0 results in the difference e(t) being directly added back on (admittedly after being redundantly multiplied by 1.0). Therefore in the very first time interval the entire PID Controller "eliminates" itself as a degenerate case.

Also noteworthy is that some PID Controllers have a "deadband" feature, to prevent over-activation: a minimum threshold below which the output does not activate. Biological neurons have this capability, and also have an "over excitation" limit as well, which presumably solves the other issue faced by PID Controllers: Integral windup - a large change which (due to Integration) amplifyingly-accumulates and overshoots.

A key insight here is that if a neuron has a tight loop back to itself, then in effect the presence of the ability to Integrate and Differentiate wrt means that a neuron bears a remarkable resemblance to a PID Controller. This begs the question: where in a neuron does the "P" (Proportionality) factor come in? It may be reasonably assumed to be part of the electropotential of the "circuit" - the loop between the neuron output and its reconnection to the input, but, fascinatingly, via a resistance: firstly the length of the axon, and secondly by the resistance introduced through the synaptic terminals. If it can be established that the input is influenced in some directly or crudely proportional way along this electrical path by the difference in the firing rate of the output measurement (which will be "gating" the input) then we have the "P" of PID Control. This would be a fascinating area of future research: a refinement and exploration of Renshaw Loops.[30]

Notably[29]:

> The rate of discharge of the Renshaw cell is broadly proportional to the rate of discharge of the associated motor neuron(s), and the rate of discharge of the motor neuron(s) is broadly inversely proportional to the rate of discharge of the Renshaw cell(s).

Already it is thus established that there exists, from the Renshaw Inhibitory neuron, a form of "P" (Proportionality) found in PID Control. The rationale as to why Nature would have the discharge rates being inversely proportional remains to be seen, and may lead to more effective design of PID Control.

With the similarity between a PID Controller effectively achieving the same task as neurons (or a pair of the same - one excitatory and one inhibitory) that have a feedback loop to (or between) themselves, it is suspected that Biological neurons will perform some sort of feedback similar to the Relay method (developed by Åström–Hägglund) to "auto-tune" and naturally develop stability. On its own this would be another fascinating area of future research.

3.8 Wave Functions

Wave Function collapse: goal is to represent 1:1 input is preserved as output, given that output is linked (looped) back as the input, and consequently has a delay.

Removing the word "Wave" as extraneous, the goal of the *Function* is to stably preserve 1:1 the relationship between input and output, despite latency.

PID Control clearly meets this definition, which then has implications back to Quantum Wave Functions given that firstly the sum of all probabilities must be 1.0 and secondly when applied to input and output there must be only one single answer. Given that *some* Wave Functions may produce 50:50 probabilities (double slit experiment) it is clearly **not** the case that all Wave Functions result in Consciousness.

The additional issue with the declaration that Quantum Mechanics probability Wave Functions can be used is that Wave Functions, by mathematical definition, have no built-in concept of time. Correct usage of Wave Functions is well-known to be "simply an expression encompassing probabilities", and further that it the *user* of that function that makes the decision to associate "meaning" to its use. The assertion that Consciousness can be expressed purely through Wave Packet Reduction is easily and unfortunately demonstrably false.

3.9 Conclusions so far

At a high level:

- many types of Advaita Vedanta knowledge map onto Boolean Logic Operations
- others describe input and output (senses)
- others describe storage (mind) or what may be assumed to be *in* that storage (facts or tenets)

Which resembles a remarkable similarity to the specification of any modern Computer, comprising:

- An Arithmetic Unit
- I/O and peripheral management

- Memory (both short and long-term)

Further aspects:

- A source of randomness and
- PID Control for real-time stabilisation and resource management

These share the same characteristics of modern Multi-processing Operating Systems such as FreeBSD, GNU/Linux, GNU/Hurd and many others.

It was also identified that at least the Renshaw Loop (one inhibitory neuron directly linked to a muscle-excitatory one) may be a type of PID Controller, and that there is the beginnings of hints that this type of "self-observing loop" that may take into consideration additional input, may be where Consciousness begins. Hankey's 2019 paper states:

> From our analysis above, critical instability seems to offer a basic solution. An infinitesimal, self-observing information loop functions like 'a self-knowing itself'. Clearly, the information flow in the loop is 'informing itself'. What does it tell it? Nothing, no measurable property (no thoughts), no qualia (no feelings) But: it provides a basis of self-reference for any information added on top of it

PID Control - which takes its own output as an input - clearly meets the criteria of being "self-observing", and also takes into consideration external information (whether a muscle has contracted or not)

So it would seem that the primary locii of Consciousness is in the self-referral loop, and that everything that surrounds it (Memory, Sensory input, outputs in the form of sound and waste byproducts in biological entities) is almost... incidental.

4 Examples meeting the Definition

This section goes through some examples that meet the definition of Consciousness *in the strict sense*. This does not make the definition correct: rather it is a two-way process. Early drafts of this paper missed out the crucial "self-referral loop" (explored below).

4.1 Operating Systems

Pre-emptive multi-tasking Operating Systems are, after over 50 years, still an active area of research for improvement. The issue is that tasks (tens to hundreds to thousands and beyond) need fair allocation, on-demand, in real-time, without resource starvation including I/O, Memory of at least three different kinds, and including negotiation for strict "locking" of resources that could become corrupted or damaged. Failure to perform adequate resource management can result in data corruption or catastrophic failure or security breach.

To complicate matters further there is frequently multiple processors available, all of which must be allocated tasks. In the case of IBM servers the number of cores can exceed 1,000.

The criteria are met as follows:

1. YES - Boolean Logic and Algebra - a given.
2. YES - differentiation and integration wrt. - achieved through software.
3. YES - read and write Memory - a given
4. YES - select inputs - also obvious.
5. YES - choose where to write - again also obvious.
6. YES - Access to a source of true randomness - absolutely necessary, for both security purposes as well as resolving contention
7. YES - Coherent Corrective regulatory real-time feedback - also essential.

The corrective feedback involves ensuring that when there are multiple tasks attempting to access the same resources, there is no oscillation between glut and famine.

Overall, it is very strange to find that a Multi-tasking Operating System could be considered "Conscious" - aware of its state, aware of its past state, and so on, and reacting to meet its purpose.

4.2 Anti-virus software

Looking at spamassassin (a FOSS product with full source code published) and its use of a service called "razor", and other services, it is clear that this very sophisticated software is capable of real-time pattern-matching (Bayesian Inference) based on information that is captured, synthesised and distributed from ongoing world-wide spam and other types of attacks.

"spamassassin" is designed to identify and cut spam. Without the additional plugins it has limited but reasonably-effective capability but needs some manual assistance in the form of "training" to recognise what is and is not "spam" (unwanted unsolicited email).

However, spamassassin's use for spam detection is not the same as Anti-virus software (although it could be used for the same), which involves additional tasks: monitoring real-time network traffic (web pages and downloads *as they are happening*), and monitoring of programs, some of which may already be running, some of which may be started without the user's knowledge.

It is quite crucial to note just how sophisticated Viruses have become: these are called "Polymorphic" viruses and they literally change themselves to evade detection. Simple pattern-matching, the mainstay of Anti-virus software, just doesn't work when the binary execution is evolving, just like a Biological virus, right before your eyes.

One very important job of Anti-virus software is to check that it has not *itself* been compromised! Additionally, one of the insidious aspects of Anti-virus software is to *modify* or replace system calls ("please read file XYZ from disk") to hide their existence! An example would be to replace the "file read" system call with one that, if the user ever tries "please read file KNOWN_VIRUS.EXE from disk", it responds "File Not Found" when the file does, in fact, clearly exist.

To overcome this problem, programs in the FOSS world such as "rkhunter" and "chkrootkit" are therefore recommended *not* to be run on a live system if possible, but to be run instead from a "known-good" secure bootable media that then double-checks the contents of a Hard Drive containing the Operating System which could have been compromised. The theory being: if the OS on that Hard Drive has been compromised, through the Virus "lying" through compromised System Calls about its existence, only by booting from *trusted* media can signs of the infection be detected.

With this background in mind, here are the criteria:

1. YES - Boolean Logic and Algebra
2. PROBABLY - differentiation and integration wrt.

3. YES - read and write Memory
4. YES - select inputs
5. YES - choose where to write
6. YES - Access to a source of true randomness
7. NOT KNOWN - Coherent Corrective regulatory real-time feedback

The latter is unfortunate and ensures that Anti-virus software may not meet the Definition. The reason is that the source code and behaviour of *Proprietary* Anti-virus software is simply not available for inspection, and FOSS Anti-virus and anti-spam software has both different roles or less capability.

It would be nice to know if Proprietary Anti-virus software had the ability to perform real-time corrective feedback in its detection of Viruses (including infections of itself) but without the source code it is just not possible to check.

4.3 Humans

The match against the Definition of Consciousness is at a low-level, particularly the Algebra. Mathematics is taught. Put another way: human babies are not born with the inherent ability to calculate pi to a thousand decimal places. Below is at the *Architectural* level of human neurophysiology (as opposed to the psychological state of mind, or the post-educational mental capacity):

1. YES - Boolean Logic and Algebra
2. YES - differentiation and integration wrt.
3. YES - read and write Memory
4. YES - select inputs
5. YES - choose where to write
6. YES - Access to a source of true randomness
7. YES - Coherent Corrective regulatory real-time feedback

All of these features are identifiably either part of the brain structure, nervous system, the neurons, or may be observed or deduced.

It really is rather unfortunate that we cannot state, categorically that all humans are Conscious - i.e self-aware. It has to be objectively acknowledged without judgement that there is sadly a large proportion of the human race who go through their lives literally by rote, responding with nothing more than "by-the-book" socially-programmed epithets their entire lives. Nothing about their existence involves Creativity, spontaneity or Creative thought. It may even be the case that such individuals displaying the "by-the-numbers" rote response characteristic never exhibit random unpredictable behaviour *of any kind*, although this would require an extensive study to prove.

When we therefore talk about Consciousnessness, it has to be borne in mind that there exists a *potential* for every human to exhibit this astonishingly-beautiful characteristic, and that even if it cannot be properly defined formally, it is a reasonable hypothesis to put forward that anyone who is themselves Consciously aware can instinctively tell if they are interacting with another human being exhibiting the same. How, exactly, when there is *no known definition of Consciousness* let alone one that is shared and agreed upon by any two given "Conscious" humans interacting with each other (even for the first time), is a startling paradox worth exploring in its own right.

Frustratingly there is not much more that can be said about the non-correlating Inference "Human=>Conscious" vs "Conscious=>Human" other than to reiterate the irony that any given Conscious being (typified by a given selected human as an example) *does not* need to know the *defined properties* of Consciousness in order to exhibit those very same properties! We (humanity) owe a massive debt of gratitude to anyone and everyone who has made even the slightest effort, successful or not, to anyone attempting throughout human history and modern times to document or study the elusive topic of Consciousness.

4.4 Gene Regulation: Epigenetics

Both Gene Regulation for disease fighting, and protein or other complex molecular assembly inside an individual Cell, follow a similar pattern of comprising multiple inter-related phase-coherent and inter-coordinated feedback loops.

In a nutshell, without introducing critical instabilities into the equation, but instead describing the *achievements* of the system:

- Disease is detected
- Anti-body production is triggered
- Effectiveness is measured
- Parallel production of anti-bodies increased or decreased
- Disease is defeated
- Anti-body production ramps down

Looking at this objectively it looks like a classic PID Control problem. Except that unlike a classic single PID (controlling a single heater for example with a single thermistor measuring boiler gas temperature), it is a massively distributed problem which requires significant and careful cohesive coordination to ensure that firstly the disease is defeated and secondly the body's resources are not exhausted due to over-production of anti-bodies.

It is not just a distributed problem as far as production is concerned, either: it is a distributed *detection* mechanism as well, *and* a distributed delivery mechanism to ensure that the anti-bodies reach the site of the disease.

It should not come as a surprise in any way to learn that evolution would come up with a distributed system, on the basis that any system that was dependent on a single point of failure would quickly be eliminated from the gene pool. Likewise we should not in any way be surprised to learn that viruses - evolutionarily - adapt, in the face of such sophistication, to utilise similar redundancy (multiple seemingly-effective-redundant genes and gene combinations) to evade defeat.

As an aside, complications arise in cases such as Epstein-Barr Virus which interferes with both Vitamin D and Zinc transport: furthermore it goes "dormant", making a full bone-marrow transplant the only effective way to eliminate it. The point being, here, that even a fully-functioning and healthy immune system may not necessarily eradicate a given disease entirely, but still achieve its "designed" (evolutionary-driven) objectives: maintain perfect health.

Epigenetics therefore has much to teach to Computer Science, especially in the area of Distributed Computing. Distributed Computing is a known-hard area of Computer Science: very few popular and useful fully-distributed software programs exist, the most well-known first highly-successful one being BitTorrent, which came under heavy political and Mafia-driven fire, touted as "criminal" in order to suppress its use. Bittorrent's primary success stems from its

search/discovery mechanism being entirely distributed, which involves Multi-dimensional (the term hypercube is used) "nearest neighbour" recursive search.

4.5 Mills Electron

Mills[32] derived formulae to describe the properties of the electron to 10 decimal places with no postulated "magic constants". Contrast this with the Standard Model which is well-known to have 27.

Of note: Mill's strong criticism of Quantum Mechanics is easily justified on observing that QM covers probabilities, and has no concept of "time". By moving everything to the Frequency Domain the mathematics becomes unnecessarily complex, but crucially the *position phase and magnitude at any given time* of components within a particle may not be considered "exact": instead "probability" is conceptually used, which immediately rules out Discrete Event-driven Simulation, Computational Fluid Dynamics, and many more *well-known* tools from being useful. Further discussion is out of scope, other than to highlight that QM is never going to successfully describe Consciousness.

The basis of Mill's theory is that a photon enters a self-standing self-phase-locked self-travelling infinite loop, and judicious use of Maxwell's Equations and Jackson show that the resultant EM Field is a perfect 2D sphere with uniform force distribution, corrective feedback, and a non-radiating Boundary Condition such that a Special Relativity Correction may be applied to turn what would otherwise be the Bohr radius into the Compton wavelength of the electron.

The crucial aspect of this Model is that, astoundingly, it *meets the definition of Consciousness*.

1. YES - Boolean Logic and Algebra - achievable through Spherical Harmonic algebra, notably superposition, and other EM interaction.
2. YES - differentiation and integration wrt. - achieved through phase-differences and the fact that the photon inside the electron has both phase and frequency
3. YES - read and write Memory - the electron is effectively a miniature magnet: very crudely put it can either be in a N-S state or a S-N state. Additionally it fits Orbital shells, *including* Fractional ones[33]
4. YES - select inputs - sort-of. this will be phase-dependent as well as involve the phase and frequency of any EM interactions.
5. YES - choose where to write - sort-of. again, the system is so simple (yet non-deterministic) that "choose" is almost the wrong word
6. YES - Access to a source of true randomness - at a high level, any external interaction will have no idea what the phase is. additionally, the electron has access to vacuum potential energy at any time.
7. YES - Coherent Corrective regulatory real-time feedback - this gets complicated and requires special consideration below.

Regarding coherence: an aspect of the Mills Model is the self-standing self-phase-locked circularly-travelling photon. The implications are that if any given looping is considered to be representative of "self" (self-state, by definition being the photon's EM phase and magnitude at any given time - I and Q in radio terms) then that loop clearly *superimposes* with all prior loops, recursively and infinitely. Further: any change at any given moment *immediately* shows up as a phase-differential between the current and old states.

This *very precisely* matches with the working definition of Consciousness: infinite self-observance. Additionally the strict Mathematical definition of Mill's theory, having no postulated magic constants of any kind, provides extremely valuable clues as to the potential Mathematics behind Consciousness. Notably: the coherence of the infinite cumulative self-superposition with respect to time.

An individual electron, being a fundamental particle, would be expected to be simple. However what is unexpected is that *within its domain* it meets the Definition of Consciousness. This requires strong emphasis: it is *not reasonable* for a tiny system to meet the same level of sophistication as a whale, dolphin, primate or crow. It is however perfectly reasonable for a system to display *appropriate* levels of "awareness" within the context of their resources.

In short: it is *not possible* to dismiss an Electron as being "Conscious" - caveat being "within the context of its resource limits".

5 Discussion

A very important question: does Consciousness arise purely from the arrangement of neurons into Cortices, or does Consciousness arise from the creation of an Electro-magnetic spectrum (measurable for example with an EEG) caused by a particular pattern that a Conscious human can "enter"? This is not in any way an easy question to ask let alone answer: up until this point, the entire paper has *assumed* that Consciousness in humans is expressed exclusively by and within the neural structure of the brain, its hormones, chemistry and its wiring.

An additional question lies in whether there is some aspect of Consciousness that requires special but previously-undiscovered aspects of either physics, Particle Theories, or the Universe itself. However we can be partly but not completely reassured that there are additional examples that meet the Definition, without requiring metaphysics or non-physical theories, but note that judgement either way is very much premature and prejudicial, and leave it as an open question for future study. Such study would need to take into account Hankey's observation[2]

> "...Sheldrake's story of 'Nkosi the African Grey parrot. 'Nkosi can accurately call out the names of images viewed by her mistress in another room (21)."

and further that

> Consciousness is associated with high levels of correlation... correlations are ... long-range criticality correlations.

First: what is Long-range Criticality?

> Long-range criticality correlations, a hallmark of self-organized criticality (SOC), describe systems where correlations extend over large distances in time and space, exhibiting power-law scaling behavior.

What comes intuitively out of this, particularly when looking at how a PID Controller ensures a 1:1 correlation between desired and actual target, in the face of both latency and noise, is the question "how may this same looping be ap-

plied to multi-valued information?" aka "a vector". To clarify: a PID Controller may be considered to be single-value self-preserving self-referral information, but what mechanism would similarly preserve *multi-valued* information? This is where Long-range Criticality potentially comes in.

Sipling Zhang and Ventra show that memory alone is sufficient to induce a phase of spatial long-range order (LRO), and it would seem that this situation is also what Hankey is describing: Long-range Criticality Correlations, which it is seen has key properties: power-law scaling behaviour and phase/frequency correlation, despite long-distance separation[10].

In other words: a "memory" in a suite of Neural Network Cortices is:

- constructed from "properties" (a qualia) not from "the data itself"
- distributed
- exhibits long-distance phase-correlations
- the rate of change of the memory is slow and
- the rate of access and use of that memory is faster than the update speed of the memory.

The construction from properties, as noted by Tononi and others, means that any given Memory is "scale-free"[6]. Gabora and Ranjan postulate that actual Memories are *reconstructed* from their associated qualia, rather than stored verbatim.

For anyone familiar with Distributed Databases (OpenLDAP for example) that is primarily designed as a read-intensive store, the above characteristics should be instantly recognised as "obviously necessary". A Slow update rate allows for distributed replication of data ("Memory") without huge design complexity, providing for much faster localised processing, *without* data corruption occurring (one Node using out-of-date information).

Hankey again notes:

> Freeman's discovery that self-organized criticality is present in EEG wave packets associated with mental cognition.

Whatever is going on, then, in Neural Networks, there appears to be evidence of at least "persistent repeated time-coincidental firing" across *multiple* neurons if not entire Cortices, and that somehow this is also a property which inherently creates not just "Memory recall", but also, perhaps through the looping (with gain=1) associated with ensuring the Memory is long-term persistent, conveys an "experience of self" alongside it.

This hints at a hypothesis where "self" is simply a suite of abstract properties, none of which have any kind of formal name, unlike "concrete" properties such as "is female" or "is blue" or "is numerically greater" or "is loud". Further, that there exists a "masking" system (known in Vector Processors as a "predicate mask") such that in any given memory-recall-and-processing, some (or even all) of the properties may be "removed from processing consideration", but *without disabling the actual recall and self-referral looping mechanism.*

```
| cortices: vector of qualia.       |
| self-properties | external ideas |
| 111111111111111 | 00000000000111 |
| transcendence (hypothesis 1):     |
| 000000000000000 | 00000000000000 |
```

Hankey proposes that Transcendence occurs when the qualia is an empty set:

```
| <=======0 |
|  <======0 |
|       .... |
|         <0 | empty set
```

An alternative proposal is that the mask instead is the empty set:

```
|  1111111  |
| <=======0 |
|  0111111  |
| <=======0 |
|  .....    |

|  0000000  |
| <=======0 | becomes equivalent to
| <       0 |
```

In other words: the occurrence of Hankey's Long-range Criticality, also known as Tonini's cause-effect lookup, is not prevented from occurring by the predicate mask happening to be zero: it goes ahead regardless but by definition is always guaranteed to produce a null result. This has some rather startling implications for the meditative "Transcendental State" of Consciousness, not least that by pure chance at any time in such a state may *any random set of mask values suddenly switch on*, inspiring what amounts to a completely new and innovative "thought / Memory" to occur - defined as "an associated set of properties aka qualia".

Furthermore it is also not hard to agree that such a completely-random (new, innovative) Thought/Memory/Property-set has a statistically-negligeable chance of occurring whilst "The Mind" is preoccupied with the processing of other "day-to-day" thoughts, as even the presence of a minimalistic qualia, a "prior experience", would interfere with and/or become part of what would otherwise be an entirely new idea (qualia). It is important to refer to Tonini's properties of IIT, notably (summarizing) that the mechanism responsible for Thought/Memory lookup may not perform two simultaneous lookups, in any way. Put another way: defining a thought as "a vector of properties aka a qualia": if "all possible associations between all possible thoughts, and the thoughts triggerable by the same" is a Directed Graph, and the associations all involve **known** properties, how can an entirely new thought, comprising properties **not in the set of prior experience**, ever be triggered? Answer: it can't. However *if* such a thought were to occur, its properties could inspire exploration to link it to "Known Experience".

5.1 Distributed Memory and Creativity

Gabora and Ranjan mention that memory is *reconstructed* rather than recalled. Further that superpositions of properties unassociated with the original memory can occur by accident.[4]

As an aside it is important to appreciate that "time" may also be considered to be a property (or a qualia in its own right), which has implications for what is considered to be different "types" ofMemory, specifically that they may not be different at all[16]. What *is* clear however as Tonini points out is that *different categories of properties* get handled by different parts of the brain. One area is dedicated to colour identification, and another involved in spatial awareness shows an expected "grid" patterning. The point ultimately being that to recall any type of Memory requires "Activation" (involve-

ment) of different categories of properties, combined to create a qualia.

If 100 percent accurate, any given qualia is a key into a distributed key-value store. However if individual qualium (a property) or subset qualia can also trigger distributed recall then the Memory acts more like a Content-Addressable Memory (CAM). Given that it looks like rudimentary processing is built-in, Conscious beings are looking more and more like an Aspex Microelectronics Array-string Processor[17]. Later Architectures (at least VASP-E thru F) had 256 bytes of CAM per Arithmetic Unit (ALU) - the VASP-F Architecture had 4096 parallel ALUs[18]

The very fact that lookups occur by properties means that accidental superposition is inevitable. A fascinating aspect of Gabora and Ranjan's work is that this gives rise to "Creativity" as well as the concept of "Insight":

> Because information is encoded in a distributed manner across ensembles of neurons interacting by way of synapses, the meaning of a representation is in part derived from the meanings of other representations that excite similar constellations of neurons.

(It is postulated that a representation is more likely to be *exclusively* derived).

> Content addressability ensures that the brain naturally brings to mind items that are similar in some ... way

by accidentally combining two or more Memories that happen to share a subset of common properties. Where the common subset is very small, a more "outlandish" - creative - thought might occur:

> She enters an associative mode of thought, and her activation function becomes flat, such that the associative structure of her memory is more widely probed.

Again this is referring to the early phases of a DHT-like algorithm where smaller qualia subsets (potentially down to a single property) are initially used as lookup keys. In short there appears to be confirmation by Gabora and Ranjan not only that there exists a type of DHT lookup in Biological Memory, but that it can accidentally - but usefully - result in "Creative insights".

5.2 Memory (cont.)

In Biological Systems there may be different types of memory (or, our understanding of how memory might arise in a Biological System is limited). Tononi discusses how IIT (Integrated Information Theory) endeavours to extrapolate the defining characteristics of Consciousness[3] From expected principles "Composition Information Integration and Exclusion" a circuit diagram is derived

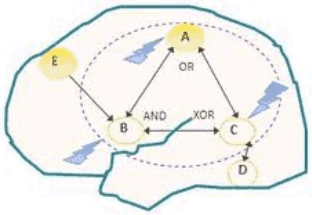

Figure 3: Tononi illustrating Intrinsic Existence

What is remarkable about this circuit is that its functionality bears a close resemblance to a bi-stable Latch

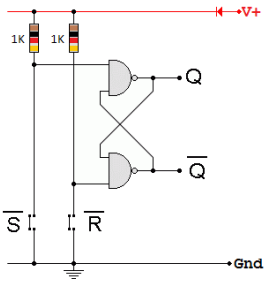

Figure 4: Bistable Latch: pair of NAND gates

Re-expressing Tononi's circuit to have inputs into both the AND and OR Gate, a quick check shows that when both inputs are zero, the state of the output from the OR gate is part of a cyclic loop between the OR and XOR. Put another way: if one input to an XOR gate is zero then the other input is alway equal to the output. Likewise with an OR gate (the difference between OR and XOR being when the second input is high, and it is when the second input is *zero* that is relevant). Therefore, linking an OR to a XOR in a cyclic loop preserves the state value, indefinitely.

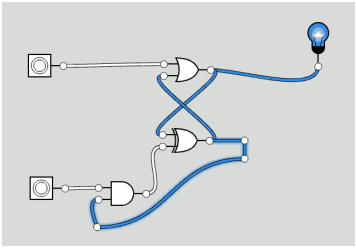

Figure 5: Simulated Tononi circuit on logic.ly

However when the AND gate's additional input is activated, it can be shown that the only supportable value output from the XOR Gate is a Zero. Likewise when the second additional input, connected to the OR gate, is activated, the loop between the OR and XOR gate has no choice but to output a One.

This behaviour is *by definition* exactly that of a Bi-stable Latch.

There is however a minor issue with the circuit *as expressed*, namely that there are potential propagation delays, depending on the distance between the three gates and the speed of signal travel, which could introduce unstable activity when the AND gate's external input drops from One to Zero. This issue can be solved by the near-symmetric addition of another OR Gate.

However much more crucial to appreciate here is that Tonini's superb work is a first-principles reverse-engineering inference, not a biological-based derivation, and as such it is possible to play with alternatives, whilst (wisely) ensuring the IIT principles noted by Tononi are preserved.

At this point, it is worth coming back to the PID Controller, or more specifically the excitatory and inhibitory pair which creates a Renshaw Loop, and noting that the pair naturally create an XOR gate. Also it is noted that an OR gate may be created by simply connecting each Axon from each input to one of the pair. An AND gate however is conceptually trickier until De Morgan's Law is brought into play - (!A AND !B) = !(A OR B). Put another way, immediately relevant given that AND can be converted to OR, which is simply "connect

Axon input": "!(!A AND !B) = (A OR B)". Noting that Gating an Inhibitor is equivalent to Suppression of an Excitor, the inversion of the contents of the brackets may be achieved by connecting the (inverted) inputs to the *other* Neuron in the pair. Only the unique properties of an XOR Gate allow this to be done.

The end-result, startlingly, may be that the Renshaw Loop *may have all of the properties described by Tononi* and also be capable of acting as a Bi-stable Latch. At which point, we note that if the gain=1 (taking into consideration that the I and D of PID are likely to be a pragmatic way of overcoming instability which a given theoretical model would not have) it also satisfied Hankey's condition as a component in a larger system displaying self-organised Criticality.

What is highly significant here is that a prior expectation that there would be separate systems for features such as Inference, Memory, Analogy etc, may all turn out to be provided by far less components than would be expected: the "system" inherently now expected to serve multiple purposes. Evolution at work, once again.

5.3 Implicit Hitting Sets

The similarity of Implicit Hitting Sets to the scope of BitTorrent Search and to "Creative insights" is remarkable:

> The standard hitting set problem involves finding a set (the hitting set) that intersects with (or "hits") every set in a given collection of sets.

An overview is given by Chandrasekaran Karp Moreno-Centeno and Vempala. [7] Rollon Larrosa and Petrova provide an innovative approach that can exceed the performance of state-of-the-art in this field (toulbar2) [9]

However where "Creative insights" effectively performs a merging of two (or more) disparate ideas, the Implicit Hitting Set finds the properties common to each of those separate and distinct ideas.

The Hitting Set Problem is remarkably similar to that of finding the Greatest Common Divisor, if a given set's members may be expressed as unique Prime Numbers and the set as the multiplication of the same. Sorenson for example describes a parallel GCD algorithm[5]

It also turns out that treating Sets as GF(2) polynomials produces an easily-paralleliseable algorithm, which has huge value given that all GF(2) arithmetic boils down to single-bit XOR operations. This would also be an additional separate relevant area of further research (with thanks and gratitude to Chaitanya Nirhali for the discussion and inspiration).

5.4 BitTorrent Search

One issue that appears intractable inside a biological system is: how are individual explicit Memories located, if they are stored in a Distributed System?In Computing, every piece of data has an Address: every sequentially-ordered piece of data has a corresponding unique address, making retrieval conceptually straightforward.

But a Biological Distributed Memory Store clearly uses a different mechanism. Put colloquially: what the heck *is* the lookup mechanism on Biological Memory? The probability is high that it is a Content-Addressable Memory, particularly given the prevalent occurrence of qualia (sets of properties).

One possible answer comes from BitTorrent's Distributed Hash Table (DHT) algorithm - the Kademlia Protocol - which is actually a robust distributed key-value store[8]

Torrent clients have a 160-bit Identifier associated with them: a string of 40 8-bit hexadecimal values. To draw in the Analogy: each 160-bit Identifier could either be a suite of 160 individual binary properties, or it could be considered to be a set of 40 8-bit properties. The search works by increasing the number of properties to be queried, effectively walking the edges of a Hypercube progressively towards the target.

Extremely important aspects of this method is that it is:

- resistant in the face of corruption,
- is fully-paralleliseable and
- only uses XOR gates.

Such remarkable simplicity would be easy for Evolution to "happen to come up with". Again therefore there is another line of potential research: exploring how lookup occurs in Distributed biological memory, and comparing it to similar Distributed Computing Algorithms.

5.5 Concept of Self

All the philosophical literature points towards the concept of awareness, awareness of awareness, and of the subject "being aware" of itself. This implies that there is a "timestamped record" of mind-state or more precisely "continuous sequence of mind-states", and further that the step-change over time, always beginning from "Now", is preserved as part of that sequence. Whether the step-change are discrete or continuous is yet to be determined, but purely for practical reasons (infinite storage capacity being highly unlikely) discrete seems more likely.

Where "concept of self" comes into play would therefore occur when a system "chose to recall any time-stamped moment including Right Now". The reassurance that comes with the anticipated recall producing an expected result would be part of "self awareness of self". "I" *literally* observe (and check, and verify, and am reassured that nothing changed) "I".

Any change (or corruption) therefore would show up as a "difference" between any two given mind-states at different times. This sounds remarkably like a built-in Checksum mechanism from Computing. Would it not be thoroughly ironic if the concept of "self" and self-awareness arose from Evolutionary pressure merely as a way to alert a Conscious being that their mind-state had become corrupted? As already noted, Quoy[12] et al observe that Error-Correction already occurs.

Fun as the idea is to entertain, it does seem unlikely given the experience of sports athletes to peacefully "witness" themselves objectively, in real-time, running their own race, in their own body. Although it does seem that such a profound "life achievement" would constitute a huge change in "self", which brings us on to consider the implications and impact of "changes in self-states".

5.6 Implications of changes in "self-state"

Any change - deltas - would (assuming only one new idea, new thought, new "way of thinking", new habit) show up as a "spike" that propagated down the chain of "sequentially-recorded self-states". If a given human, for example, was simply not used to such delta spikes occuring, the probability of fear or other psychological resistance to such change would be judged (by the "Experiencer") to be too high to accept.

In other words, astoundingly there may be a *mathematical* reason why people resist behavioural change and new or challenging ideas!

Coming back, then, to the Athlete who realises, towards the end of a race, that a life-long dream is about to be fulfilled, to a concept known as "Witnessing". Soldiers hit by sudden traumatic incidents also have described themselves "Witnessing" the events they are experiencing. Anyone who has been psychologically tortured, who naturally reacts in "disbelief" at the way they have been treated, is also known to go into what is termed "Dissociative Shock" but the phenomenon bears a remarkable similarity to Witnessing.

The basis: an entirely new "self-state" is emerging that is radically different from their previous experience. Whilst the new *and old* self-states are still being strongly held (only a few time-deltas have passed), the cognitive dissonance would naturally trigger the strange effect of "being an observer of ones-self" as the Athlete can easily recall the prior Memory of their former "self". But unlike an individual experiencing fear at any given radical change in "self", the Athlete is in a state of euphoria at their achievement (about to be fulfilled) and so does not resist the change as badly: instead they may display significant emotion at the successful transition to their new "self-state" of achieving life-long goals.

This has implications for Schizophrenia, which is likely to be the denial and "blocking away" of events so traumatic that the result of such blocking creates two entirely separate and mutually-exclusive "mind-states". This is not to be confused with bi-polar disorder which is where *one* mind-state experiences mutually-exclusive severe depression or hyper-stimulation. Further research is needed here.

5.7 Transcendence

The implications for a Transcendent State of Consciousness are puzzling. Where normally, a Memory Recall would trigger a qualia, it is established that the result of being in a Transcendent State is that the request - which still proceeds - is for *an empty qualia*: no properties activated as part of that request, and consequently the answer *should* also (always) come back "No results found".

However when this action is spontaneously and continuously triggered, including recalling (or more precisely *not* recalling) past "mind-states", what does that even mean?

Is it more likely, therefore, that the Transcendent State recalls only a qualia of *self*-properties, not encumbered by the additional "lookup" work of day-to-day thought as part of that same lookup? Put more clearly:

- A normal Recall involves a qualia comprising a combined set of "self" properties mixed with "the activity" properties.
- A Transcendent Recall (1) is hypothesised to involve an entirely blank lookup (full masking).
- A Transcendent Recall (2) is hypothesised to involve just a qualia lookup using "self" properties only.

```
| cortices: vector of qualia.      |
| self-properties | external ideas |
| 111111111111111 | 00000000000111 |
| transcendence (hypothesis 1):    |
| 000000000000000 | 00000000000000 | empty
| transcendence (hypothesis 2):    |
| 000000000000000 | 00000000000000 | zero mask
| 111111111111111 | 00000000000000 | properties
```

However it is important to note that it is hard to tell which (self properties are looked up, or are not looked up), in a Transcendent State. Further study is needed.

5.8 Simple PID Controller

A Simple PID Controller also barely meets the Definition of Consciousness, but in a strict purposeful mathematical way. The reason is that its construction is very similar to the Mills Electron Model, but crucially it is easier to understand.

1. YES - Boolean Logic and Algebra - achievable by definition.
2. YES - differentiation and integration wrt. - again by definition, the I and D of "PID"
3. YES - read and write Memory - an internal variable is needed: the desired target value, from which the "difference" is computed and fed into each of the three proportional contributions.
4. YES - select inputs - if the P I and D constants may change over time (such as in a biological neuron e.g Renshaw Pair) then input may be "selected" or at least "used in calculations".
5. YES - choose where to write - again sort-of, by varying P I and D.
6. YES - Access to a source of true randomness - PID Controllers *by definition* are designed to take into account randomness, and generate corrective feedback.
7. YES - Coherent Corrective regulatory real-time feedback - again by definition a PID Controller is *designed* around corrective feedback.

It can be seen that in the strictest sense a PID Controller can be deemed "conscious", given that "self" state and crucially the *difference* between desired and actual target value, is infinitely looped-in to the output. Also fascinating to note that the design of the "traditional" PID Controller has the same "instability" as humans if presented with too much information (too big a state-change). This is down to the Integral part creating too large a cumulative contribution at start-up, sending the system into a wildly-oscillating state.

It is remarkable that Engineering Design analytically solved the same problem as Biological Systems, and, again, the PID Controller provides key Mathematical insights.

5.9 Clustered Simple Biological PID

This needs further research. Hankey and others state that simple O(1) critical instability systems (a critical instability with only one feedback loop) is not possible to demonstrate the characteristics of Consciousness. However given that a single PID Controller can in fact do precisely that means that something else is going on.

What *might* be occuring is that the O(1) critical loop as described by Hankey and others may have an I=0, D=0 or both: i.e. the feedback (gain=1) is only based on the D of PID Control. Such an O(1) - single PID controller - would definitely display instability, because in effect with gain=1 and no I or D it is as if the PID Controller was not present at all.

Hankey points out that in for example Epigenetics, a biological system will ramp up more "production" of e.g. an anti-body in response to viral infection. Considering the Distributed Production as a whole, comprising *multiple* O(1) simple PID controllers that have, *on their own* no I or D, the

gradual activation of *multiple* such simple PID controllers *in effect* introduces an I (Integration) term into the *overall* system. The reason is that the activation, performed over time in response to a single (global) measurement of the quantity of virus detected, *accumulates* (activates) more and more antibody Producers.

Advantages of this approach - from an evolutionary perspective - should be very clear: there exists no critical single point of failure. The entire system is robust, being distributed, and as a side-effect the individual Production Units need not themselves have a complex Integration function built-in: Integration is instead achieved by the system as a whole using much simpler components. Further study is needed to identify whether the measurement system is also distributed: given that robustness and redundancy is key to the success induced by Evolutionary pressure, it is highly suspected that this will also be the case.

5.10 Grand Unified Field Theories

It is believed reasonable and rational to state that any Consciousness entity (regardless of scale) must obey the Laws of Physics. To whit: it is only the Human "belief system", as a form of prejudicial self-judgement that, if stretched beyond breaking point, prevents mathematical exploration and application of creative insights. The implications for HEP: any correct Model of how the Universe actually operates *must* allow for Consciousness and all known-phenomenon (whether subjective or objective) to be part of the same. Put more clearly with an example: it is absurd to have a Grand Unified Field Theory that cannot account for either telepathy or for the soul.

6 Conclusion

This paper set out to explore Consciousness: to derive and demonstrate a working Definition, based on observed historical discussion of the Properties of Consciousness. Tononi terms this pursuit "Integrated Information Theory" and makes highly significant headway but does not go so far as to create a *Definition*. This paper succeeds in both creating a Definition, outlines examples which meet the Definition, but then corrects the same whenever an anomaly is met.

The biggest surprise was to find that Consciousness - being aware of one's self - is actually rather mundane. It *literally* requires a definition of "self", and the ability to "observe" that self, and also to be able to make "observations" against repeated time-coherent snapshots. In the case of the Mills Electron it is circular-light-looping obeying Maxwell's Equations to create infintely-circling self-standing self-interacting self-observing travelling waves. In the case of a human (with a biological neural network) it is the continuous snapshotting of "mind-state" qualia - properties of the self - and comparison of the same to ensure that they are not corrupted. This sounds terribly mundane but it is the *implications* that are profound and need further exploration.

Also a significant surprise was how similar a human biological neural network is, *in functional terms*, to a Vector Processor running a standard Multi-Threading pre-emptive Operating System, and running mapreduce style Distributed Processing. All of the pieces of the puzzle are there: if the Definition is correct then Modern Computing is not that far off from achieving "Conscious Computing".

A large quantity of further research also came out of this paper:

1. Exploring self-awareness
2. Exploration of Renshaw Loops.
3. Lessons that PID auto-tuning can learn from biology
4. Implicit Hitting Sets algorithms
5. Why do human babies take longer than animals to learn to stand?
6. "lookup" in Distributed biological memory, compared to Distributed Computing.
7. Do unconscious (socially "by-rote-reactive") humans exhibit less random (less creative) behaviour?
8. How do "Conscious" humans instinctively recognise each other's "Consciousness"?
9. Clusters of Simple Biological PID Controllers.
10. What are the implications for *Collective* Consciousness, if Consciousness is "Defined"?
11. Are there implications for Schizophrenia and Bi-polar Disorder? Can a Definition of Consciousness help?
12. Does Evolution meet the Definition of Consciousness?
13. Is Transcendence an "empty qualia" or "an empty qualia mask"?
14. Epigenetics: is the "measurement" also distributed?
15. Can a Definition of Consciousness help solve Supply Chain Traffic Optimisation?
16. Metaphysical aspects: the relationship to, and role of, "the soul"
17. Implications for High-Energy Physics Models

Finally, partly as a joke but also as an insight: a Scientific Paper is by convention written third person impersonal. On the topic of Consciousness, it is particularly ironic to write entirely in such a style, when the whole purpose is to explore "subjective self-awareness". Without breaking this convention it can be stated that:

> As the author *I* feel an overwhelming urge to say that *I* really enjoyed writing this paper but at 12,000 words *I* am relieved it is done.

7 Acknowledgements

With thanks and gratitude to Prof. Alex Hankey, Dr A Akilbasha, and to Chaitanya Nirhali for the discussion and insights, and special thanks to anyone throughout Ancient and Modern times for taking an interest to explore this deeply fascinating topic.

References

[1] Alex Hankey, 2019, J. Phys.: Conf. Ser. 1251 012019

[2] Alex Hankey, Cosmos and History: The Journal of Natural and Social Philosophy, vol. 10 no. 1 2014 p237-250, Complexity Biology-based Information structures can explain subjectivity objective reduction of wave packets and non-computability

[3] Giulio Tononi (2015), Scholarpedia, 10(1):4164. doi:10.4249/scholarpedia.4164

http://www.scholarpedia.org/article/Integrated_Information_Theory

[4] Ranjan, A., Gabora, L. (2013). How insight emerges in a distributed, content-addressable memory. In A. Bristol, O. Vartanian, J. Kaufman (Eds.), The neuroscience of creativity (pp. 19-43). doi:10.7551/mitpress/9780262019583.003.0002

https://arxiv.org/pdf/1106.3600

[5] Jonathan P. Sorenson, A Randomized Sublinear Time Parallel GCD Algorithm for the EREW PRAM,

https://arxiv.org/pdf/0907.0718

[6] Klaus Linkenkaer-Hansen, Vadim V Nikouline, J Matias Palva, Risto J Ilmoniemi, Long-Range Temporal Correlations and Scaling Behavior in Human Brain Oscillations, J Neurosci. 2001 Feb 15;21(4):1370–1377. doi: 10.1523/JNEUROSCI.21-04-01370.2001, PMCID: PMC6762238 PMID: 11160408

https://pmc.ncbi.nlm.nih.gov/articles/PMC6762238/#scale-free

[7] https://faculty.cc.gatech.edu/~vempala/papers/implicit.pdf

[8] https://pdos.csail.mit.edu/~petar/papers/maymounkov-kademlia-lncs.pdf

[9] Emma Rollon, Javier Larrosa, Aleksandra Petrova, Anytime Cooperative Implicit Hitting Set Solving

https://arxiv.org/pdf/2501.07896

[10] C. Sipling, Y.-H. Zhang, M. Di Ventra, Memory-induced long-range order in dynamical systems,

https://arxiv.org/abs/2405.06834

[11] https://m.youtube.com/watch?v=pGp43w-391M

[12] Alexandre Pitti https://orcid.org/0000-0002-6541-578X, Claudio Weidmann https://orcid.org/0000-0002-7073-9981, and Mathias Quoy, Digital computing through randomness and order in neural networks, 119 (33) e2115335119, https://doi.org/10.1073/pnas.2115335119

https://www.pnas.org/doi/10.1073/pnas.2115335119

[13] https://plato.stanford.edu/entries/reasoning-analogy/

[14] https://github.com/susantabiswas/Word-Analogy/blob/master/Word%20Analogy.py

[15] Jinyung Hong, Theodore P. Pavlic, An Insect-Inspired Randomly Weighted Neural Network with Random Fourier Features For Neuro-Symbolic Relational Learning,

https://ceur-ws.org/Vol-2986/paper10.pdf

[16] https://en.wikipedia.org/wiki/Explicit_memory#Types

[17] Rohrback and Vesztergombi, Intelligent Trigger by Massively Parallel Processors for High Energy Physics, 3rd Int. Conf. Accel. Large Exp. Phys. Control Syst. ICALEPCS1991, Tsukuba, Japan JACoW Publishing ISBN: 978-3-95450-254-7 ISSN: 2226-0358 doi:10.18429/JACoW-ICALEPCS1991-S09DPP03

https://inspirehep.net/files/525f69f3ad9400ae51d4d5a5253c4c53

[18] Argy Krikelis, I.P. Jalowiecki, D. Bean, R. Bishop, A Programmable Processor with 4096 Processing Units for Media Applications,

https://www.researchgate.net/publication/2915463

[19] https://memgraph.com/blog/cosine-similarity-python-scikit-learn

[20] https://en.m.wikipedia.org/wiki/Exportation_(logic)

[21] https://en.m.wikipedia.org/wiki/Propositional_calculus

[22] https://en.m.wikipedia.org/wiki/Nyaya

[23] https://en.m.wikipedia.org/wiki/Phenomenology_(philosophy)

[24] https://en.m.wikipedia.org/wiki/Infinite_monkey_theorem

[25] https://en.m.wikipedia.org/wiki/Kashmir_Shaivism

[26] https://m.youtube.com/watch?v=0OnL_AX6Q9g

[27] https://en.m.wikipedia.org/wiki/Turing_test

[28] https://en.m.wikipedia.org/wiki/Proportional%E2%80%93integral%E2%80%93derivative_controller

[29] https://en.m.wikipedia.org/wiki/Renshaw_cell

[30] http://digfir-published.macmillanusa.com/kolbintro5e/asset/img_ch5/kolb5e_fig05_09.html

[31] Bioelectromagnetic and subtle energy medicine: the interface between mind and matter, Paul J Rosch, Ann N Y Acad Sci. 2009 Aug:1172:297-311., PMID: 19735252 DOI: 10.1111/j.1749-6632.2009.04535.x,

https://pubmed.ncbi.nlm.nih.gov/19735252/

[32] The Grand Unified Theory of Classical Physics, Dr. Randell L. Mills,

https://brilliantlightpower.com/book-download-and-streaming/

[33] Spherical harmonics for fractional quantum numbers l=1/n, Lv Qingzhang, (School of Chemistry and Chemical Engineering, Henan Normal University, Henan, P. R. China 453007),

https://arxiv.org/abs/2301.12433