Independent Core Observer Model (ICOM) Theory of Consciousness as Implemented in the ICOM Cognitive Architecture and the Associated Consciousness Measures

David J Kelley[1] and Dr. Mathew A. Twymon[2]

[1] AGI Laboratory – David@ArtificialGeneralIntelligenceInc.com
[2] AGI Laboratory – Amon@ArtificialGeneralIntelligenceInc.com

Abstract. This paper articulates the fundamental theory of consciousness used in the Independent Core Observer Model (ICOM) research program and the consciousness measures as applied to ICOM systems and their uses in context including defining of the basic assumptions for the ICOM Theory of Consciousness (ICOMTC) and associated related consciousness theories (CTM, IIT, GWT etc.) that the ICOMTC is built upon. The paper defines the contextual experience of ICOM based systems in terms of a given instances subjective experience as objectively measured and the qualitative measure of Qualia in ICOM based systems

Keywords: Theory of Consciousness, GWT, ICOM, AGI, Cognitive Architecture, CTM, IIT

1 Introduction

Designing an Artificial General Intelligence (AGI) System includes a number of complex problems including defining elements that are not agreed upon to even have a foundation to build such a design on in the first place. In the research program for the cognitive architecture for AGI called the "Independent Core Observer Model" to be able to frame tests and measures we needed to address definitions of consciousness including baking out our own version of a theory of consciousness [20] previous to this work and the process of measuring that systems consciousness which is the subject of this paper. Keep in mind these tests or measures are focused on measuring 'consciousness' not on hardware capacity or other technical measures.

The Independent Core Observer Model Theory of Consciousness is partially built on the Computational Theory of Mind [1]where one of the core issues with research into artificial general intelligence (AGI) is the absence of objective measurements as they are ambiguous given the lack of agreed upon objective measures of consciousness [2] To continue serious work in the field we need to be able to measure consciousness in a consistent way that is not presupposing different theories of the nature of conscious-

ness [3] and further not dependent on various ways of measuring biological systems [4] but focused on the elements of a conscious mind in the abstract. With the more nebulous Computational Theory of Mind, research into the human brain does show some underlying evidence to the same.

2 Our Assumptions

We made a number of assumptions to provide an experimental reference point when designing AGI cognitive architecture [20]. First, Qualia is the subjective experience that can be measured external to the system if the mind in question is operating under known parameters. [6] [19]

Humans are not able to make logical decisions. Looking at the neuroscience behind decisions we already can prove that humans make decisions based on how they feel [7] and not based on logic [27][13].

Subjective experience can be measured and understood. The traditional view that the subjective nature of experience [8] is purely subjective is rejected as a matter of principle in this paper. Consciousness, even by scientists in the field, frequently consider it the realm of "ontology and therefore philosophy and religion" [9] our assumption is that this is false.

Consciousness can be measured. "Despite this enormous commitment to the study of consciousness on the part of cognitive scientist covering philosophical, psychological, neuroscientific and modelling approaches, as of now no stable models or strategies for the adequate study of consciousness have emerged." [10] That is until now with the ICOMTC. We also believe that we can measure consciousness regarding task accuracy and awareness as a function of stimulus intensity [11] that applies to brain neurochemistry as much as the subjective experience from the point of view of systems like ICOM.

We have a concrete definition of 'Subjective' as a concept. 'Subjective' then is defined as the relative experience of a conscious point of view that can only be measured objectively only from outside the system where the system in question experiences things 'subjectively' as they relate to that systems internal emotional context.

Consciousness is a system that exhibits the degrees or elements of the Porter method for measuring consciousness regarding its internal subjective experience. [5] While the dictionary might define consciousness subjectively in terms of being awake or aware of one's surroundings [12] this is a subjective definition, and we need an 'objective' one to measure and thus the point we are assuming for the context of the ICOM theory of mind and the ICOM research altogether.

3 Basis for Design of the ICOM Theory of Consciousness

The ICOM or Independent Core Observer Model Theory of Consciousness (ICOMTC) is based on the Computational Theory of Mind [1] which is de-fined as:

According to CCTM, the mind is a computational system similar in important respects to a Turing machine, and core mental processes (e.g., reasoning, decision-making, and problem-solving) are computations similar in important respects to computations executed by a Turing machine [1] - which can have numerous variations.

. An instance of an ICOM system would be a variational instance of CCTM. In addition to that, the ICOM Theory of Consciousness or ICOMTC also borrows from the Integrated Information Theory [14]. CCTM does not give us a complete basis for developing ICOM systems and includes elements of Integrated Information Theory as well as CCTM.

Integrated information theory or IIT, approaches the relationship between consciousness and its physical substrate by first identifying the fundamental proper-ties of experience itself: existence, composition, information, integration, and exclusion. IIT then postulates that the physical substrate of consciousness must satisfy three key points or 'Axioms' [14].

ICOMTC also borrows from Global Workspace theory in that things move through the system and only when things reach a certain point is that bit of 'thought' or 'context' raised to the level of the conscious mind. [15] CCTM, IIT and Global Workspace all exist more or less in ICOMTC where ICOMTC based systems exhibit all the elements of all of these theories to some degree but it is also substrate independent in that ICOMTC is not an attempt to produce the same kind of system as the biological substrate of the human brain or do anything that requires that kind of hardware nor is it tied to current computer architecture either other than any Turing Machine [20]. Any Turing Machine in theory would be able to run an ICOMTC based system given enough processing time.

4 The Independent Core Observer Model Theory of Consciousness (ICOMTC)

At a very high level, ICOM as a cognitive architecture [20] works by streaming data and context processed by the underlying system (the observer) and based on emotional needs and interests and other factors in the system, these are weeded out until only a certain amount are processed, or 'experienced' in the 'core' (or global workspace) which holds emotional models based on Plutchik's [18] work. These elements of the core exist for both conscious and subconscious emotional landscapes of the system where the context that is 'experienced' from the standpoint of the system is the only 'experiences' that the conscious system is aware of. In this way, only the differential experience matters and the system, for example, doesn't understand a word as much as it feels the emotional context of the word as it relates to underlying context. It is the emotional valences associated with things that the system then selects things to think emotionally about. The system select's actions based on how they improve the experiences of those emotional valences and in this way the system may choose to do something logical based on how it feels about it, or it could just as easily pick

something else for no other reason than it feels a bit better about it. In this way, the system does not have direct access to those emotional values nor is a direct function of the algorithms, but it is an abstraction of the system created by the core that can be considered emotionally conscious or self-aware being sapient and sentient in the abstract.

5 Subjective Experience in the ICOM Cognitive Architecture

How do we then look at a system that experiences emotional, subjective experience objectively? The following set notation shows us a simple logical implementation of the last climb of "a thought" as it makes its rise from the depths of the system to the awareness of the conscious, self-aware parts of the system.

$$\forall\{E1, E3, \dots, E72\} \in Conscious\ , E1 = Emotion1, E2 = Emotion2, \dots, E72 = Emotions72$$

;

$$\forall\{AE1, E3, \dots, E72\} \in Subconscious\ , E1 = Emotion1, E2 = Emotion2, \dots, E72 = Emotions72$$

;

$$\forall NewContext = f(\sum Inputs)\ or\ f(MemoryStack_n)\ ,$$

$$\forall NewContext = fNeeds(NewContext)\ ,$$

$$\forall\{f\} \in ConsciousRules\ \wedge\forall\{E1, E3, \dots, E72\} \in Conscious\ , A = f(A \in Conscious\ , \{E1, E3, \dots, E72\} \in NewContext\ ), B = f(B \in Conscious\ , \{E1, E3, \dots, E72\} \in NewContext\ ), \dots, D = f(D \in Conscious\ , \{E1, E3, \dots, E72\} \in NewContext\ )$$

;

$$\forall\{f\} \in SubconsciousRules\ \wedge\forall\{E1, E3, \dots, E72\} \in Subconscious\ , A = f(A \in Subconscious, \{E1, E3, \dots, E72\} \in NewContext\ ), B = f(B \in Subconscious\ , \{A, B, C, D\} \in NewContext\ ), \dots, D = f(D \in Subconscious\ , \{E1, E3, \dots, E72\} \in NewContext\ )$$

;

$$\forall\{f\} \in SubconsciousRules\ \wedge\forall\{E1, E3, \dots, E72\} \in Concious\ , A = f(A \in Subconscious, \{E1, E3, \dots, E72\} \in NewContext\ ), B = f(B \in Subconscious\ , \{E1, E3, \dots, E72\} \in NewContext\ ), \dots, D = f(D \in Subconscious\ , \{E1, E3, \dots, E72\} \in NewContext\ )$$

;

$$\forall\{f\} \in NewContextRules\ \wedge\forall\{E1, E3, \dots, E72\} \in NewContext\ , A = f(A \in NewContext\ , \{E1, E3, \dots, E72\} \in Conscious\ ), B = f(B \in NewContext\ , \{E1, E3, \dots, E72\} \in Conscious\ ), \dots, D = f(D \in NewContext\ , \{E1, E3, \dots, E72\} \in Conscious\ )$$

;

$$\forall Action = fObserver(NewContext)\ ;$$

$$\forall\{N\} \in MemoryStack_n\ = f(\text{NewContext}, MemoryStack)$$

Fig. 1. Core Logic Notation.

First, let us walk through the execution of this logic. Coming into the system we already have context data decomposition, sensory input, also related data from memory that may be of emotional interest but for the purposes of one 'thought' let's say it's one bit of context meaning an emotionally related context tree related to something that the system has sensed externally. This will be represented by 'Inputs.' At this point, we have already passed the point of that 'context' being raised to the global workspace. Figure 1 essentially is one cycle of the core considering what is in the global workspace or 'core' of ICOM. In Figure 1 we first see that we have two sets or collections of emotional models represented by the two sets defined in the first two rows, then we have the input new context placed in the 'NewContext' set. We apply the 'Needs' function that applies a matrix set of rules such as the technical requirements of the system to other wants and needs based on the systems hierarchy of needs and current environmental conditions. At this point, we look at how this thought applies conscious emotional rules in the function 'ConsciousRules' and then how that manipulates the current conscious emotional landscape. We say 'land-scape' because it is not a single emotion but a complex set of almost infinite combi-nations consciously and subconsciously that the system experiences.

In like manner, the system applies subconscious rules to the subconscious states and the subconscious rules to the conscious states and finally those states as they apply to the new context where in all cases it is only in the abstract from this states that the system experiences anything. Meaning the system is using the abstracted states to represent that emotional landscape in how things affect all of those emotional states and related context finally being passed to the observer for action if that 'NewContext' contained an action. In this way, the system doesn't even deal with the complexity of its actions as much as the system will do them if the system felt like it and knows how; where as numerous cycles might have to execute in the core for it to perform a new task, meaning it will have to think a lot more about something it doesn't know how to do. After that context is posted back to the observer (the more complex part of the system in ICOM), then it is placed back into context memory, and in this way, we see the rich set of the emotional landscapes of the system can model and execute.

Interestingly enough, in current ICOM research there are indications that this sort of system is perfectly capable of becoming mentally ill and even forgetful if hardware starts to limit operations, where as the only way to optimize for the execution environment would be to place memory limits and based on the node map memory models this would be the only way to continue optimal execution given certain limits.

A better way to think of ICOMTC is that not a single element of the system is conscious or self-aware to any level, it is the 'interactions' between the parts that together those interactions become aware abstractly, and it is through the underlying process articulated in Figure 1 that is then measured in terms of consciousness via the various methods as well as direct instrumentation of the system to measure 'qualia' for example.

6 Measuring Qualia

In ICOMTC qualia can be objectively measured through the differential between the conscious emotional landscape of the system represented by a Plutchik model along with the subconscious model and the model of the irreducible set of any given con-text experienced by the system and the emotional model created that represents that specific 'contextual' experience. In the ICOMTC the qualia is that differential between the state and effective one emotional structure that represents that current context and how the system applies choices is then based on that and the numerous underlying factors that affect the construction and choices based on specific con-texts. Now by its nature the system can't self-reflect directly on those values but is an abstraction of that process in the global 'work space' that effectively is created by the underlying operation. We can of course measure this 'qualia' of the system but the system can't do it directly from its standpoint. In the research already done for ICOM we can see that ICOMTC system doesn't really have free will but it would appear that way from the systems standpoint and experience the illusion of free will much the way humans do.

As stated, qualia (in ICOM) then can be measured. Referring back to figure one we can use two values or sets from that set of operations and preform a 'qualia' measurement like this based on those values:

$$\forall\{E1, E3, \dots, E72\} \in ConsciousBefore\ , E1 = Emotion1, E2 = Emotion2, \dots, E72 = Emotions72$$

;

$$\forall\{AE1, E3, \dots, E72\} \in SubconsciousBefore\ , E1 = Emotion1, E2 = Emotion2, \dots, E72 = Emotions72$$

;

$$\forall\ ConsciousAfter\ = fCoreProcess(ConsciousBefore\ )\ ,$$
$$\forall\ SubconsciousAfter\ = fCoreProcessS(SubconsciousBefore\ )\ ,$$

$$\forall\ \text{ConsciousQualia[i]}\ = ConsciousBefore[i] - ConsciousAfter\ [i]\ ,$$
$$\forall\ \text{SubconsciousQualia[i]}\ = SubconsciousBefore[i] - SubconsciousAfter[i]$$

,

Fig. 2. Core Logic Notation.

In this case we are computing qualia by taking the sets that represent the cur-rent emotional landscape of the system and a conscious and subconscious level and computing the difference matching sets where a set is a Plutchik model with 8 floating point values. We subtract the current state from the previous state giving us the Plutchik representation of the subjective emotional differential experienced by the system. This really gives you the numbers in terms of 'sets' that show how a specific element of 'context' that managed to make it to the global work space is 'experienced' or rather the effective of that experience. We actually have to calculate this after the

fact external to the system as it is not actually computed in the real process (noted in figure 1) and there is not a 'direct' method in ICOM to surface an objective measure of qualia to the system without a complete abstraction but we can compute it external and use it for analysis.

The Independent Core Observer Model Theory of Consciousness (ICOMTC) addresses key issues with being able to measure physical and objective details as well as the subjective experience of the system (known as qualia) including mapping complex emotional structures, as seen in previously published research related to ICOM cognitive architecture [20]. It is in our ability to measure, that we have the ability to test additional theories and make changes to the system as it currently operates. Slowly we increasingly see a system that can make decisions that are illogical and emotionally charged yet objectively measurable [16] and it is in this space that true artificial general intelligence that will work 'logically' similar to the human mind that we hope to see success. ICOMTC allows us to model objectively subjective experience in an operating software system that is or can be made self-aware.

7 Measuring Conscious Systems

There are two types of test types that are considered for use in the ICOM program designed around measuring and testing outside of the qualia analytics that are external measures. Keep in mind qualia under the Yampolskiy method as noted below is a different measure then the previous section. These tests us allow us to measure somewhat more subjective tasks based on our behavior of the system to further bake additional research. In both cases, these tests can be applied across various potentialy 'conscious' systems and humans giving us a frame of reference for comparison which we lack using the Qualia method form the previous sections. The two test types are:

7.1 Qualitative Intelligence Tests

Intelligence Quotient (IQ) tests -. are tests designed to measure 'intelligence' in humans [24] where we are using short versions to assess only relative trends or the potential for further study, whereas given the expected sample size results will not be statistically valid, nor accurate other than at a very general level, which is believed to be enough to determine if the line of re-search is worth going down. Of these tests, two types will be used in the study, one a derivative of the Raven Matrices Test [22] designed to be culturally agnostic, and the Wechsler Adult Intelligence Scale (WAIC)[23] Test which is more traditional. Lastly falling into the category of WAIC there is a baseline full Serebriakoff MENSA test that we can apply to compare and contrast scores between the two base lines tests. [17]

Collective Intelligence (CI) Test. – we would like to use this test, however the information for executing this test is not publicly accessible and reaching out to the researchers that created this test has produced no response thus far. [21]

7.2 Extended Meta Data and Subjective Tests

A number of tests or measures will be collected, more oriented towards analysis for further study, primarily around correlative purposes. None of these tests may be used outside of as possible illustrative examples, without being statistically valid given the lack rigor or subjective nature of these measures.

The Turing Test –. this test is not considered quantifiable and there is debate over whether this measure tells us anything of value, however a test regimen for this has been completed and can be used for subjective analysis only.

The Porter Method –. This appears to be a qualitative test, but individual question measures are entirely subjective and therefore the test lacks the level of qualitativeness to be valid without a pool of historical values to measure against at the very least. This test provides some value in meeting colloquial standards of consciousness and is more comprehensive then some of the other tests albeit subjective it is at least the attempt at being a comprehensive measure of consciousness. [5]

The Yampolskiy Qualia Test –. is a subjective measure of a subjective 'thing' and therefore not a qualitative measure, however we have built a regimen based on this when looking at qualia as measured in the previous examples. In theory this only tests for presence of Qualia in human like subjects, passing this test does not mean that a subject does not experience qualia in the sense of this paper, just that it was not detected. This means that subjects may show signs of qualia, or not, but the test only would show the presence of not the absence of qualia. [26]

8 Conclusions

There are numerous potential methods for measure consciousness in ICOM or other AGI systems but there is a lot of discord in terms of a foundation to build on. For ICOM to move forward as a cognitive architecture it is important that we build that foundation. The tests listed in this paper are the ones currently being used or being considered. There are a few others including mental health states analysis tests for example but these tests are more subjective and not helpful in the current research. We have found that by building on the basis articulated here that we at least can move the research forward for the time being.

Conclusions based on the this work at least have a frame of reference even if later proved false then we at least know that and can start over. Ideally this foundation gives us a deeper more rigorous program moving forward in terms of AGI Cognitive Architecture research into conscious systems with an eye towards safety [25].

References

1. Rescorla, M.; The Computational Theory of Mind; Stanford University 16 Oct 2016; http://plato.stanford.edu/entries/computational-mind/

2. Seth, A.; Theories and measures of consciousness develop together; Elvsevier/Science Direct; University of Sussex
3. Dienes, Z; Seth, A.; The conscious and unconscious; University of Sussex; 2012
4. Dienes, Z; Seth, A.; Measuring any conscious content versus measuring the relevant conscious content: Comment on Sandberg et a.; Elsevier/ScienceDirect; University of Sussex
5. Porter III, H.; A Methodology for the Assessment of AI Consciousness; Portland State University Portland Or Proceedings of the 9th Conference on Artificial General Intelligence;
6. Gregory; "Qualia: What it is like to have an experience; NYU; 2004 https://www.nyu.edu/gsas/dept/philo/faculty/block/papers/qualiagregory.pdf
7. Camp, Jim; Decisions Are Emotional, Not Logical: The Neuroscience behind Decision Making; 2016 http://bigthink.com/experts-corner/decisions-are-emotional-not-logical-the-neuroscience-behind-decision-making
8. Leahu, L.; Schwenk, S.; Sengers, P.; Subjective Objectivity: Negotiating Emotional Meaning; Cornell University; http://www.cs.cornell.edu/~lleahu/DISBIO.pdf
9. Kurzweil, R.; The Law of Accelerating Returns; Mar 2001; http://www.kurzweilai.net/the-law-of-accelerating-returns
10. Overgaard, M.; Measuring Consciousness - Bridging the mind-brain gap; Hammel Neurocenter Research Unit; 2010
11. Sandberg, K; Bibby, B; Timmermans, B; Cleeremans, A.; Overgaard, M.; Consciousness and Cognition - Measuring Consciousness: Task accuracy and awareness as sigmoid functions of stimulus duration; Else-vier/ScienceDirect
12. Merriam-Webster - Definition of Consciousness by Merriam-Webster - https://www.merriam-webster.com/dictionary/consciousness
13. Wikipedia Foundation; Turing Machine; 2017; https://en.wikipedia.org/wiki/Turing_machine
14. Tononi, G.; Albantakis, L.; Masafumi, O.; From the Phenomenology to the Mechanisms of Consciousness: Integrated Information Theory 3.0; 8 MAY 14; Computational Biology http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1003588
15. Baars, B.; Katherine, M; Global Workspace; 28 NOV 2016; UCLA http://cogweb.ucla.edu/CogSci/GWorkspace.html
16. Chalmers, D.; Facing Up to the Problem of Conscious-ness; University of Arizona 1995
17. Serebriakoff, V; "Self-Scoring IQ Tests;" Sterling/London; 1968, 1988, 1996; ISBN 978-0-7607-0164-5
18. Norwood, G.; Deeper Mind 9. Emotions - The Plutchik Model of Emotions; http://www.deepermind.com/02clarty.htm 403 (2/20/02016)
19. Kelley, D.; Critical Nature of Emotions in Artificial General Intelligence; IEET 2016; https://ieet.org/index.php/IEET2/more/Kelley20160923
20. Kelley, D.; "The Independent Core Observer Model Computational Theory of Consciousness and Mathematical model for Subjective Experience"; ITSC 2018
21. Engel, D.; Woolley, A.; Chabris, C.; Takahashi, M.; Aggarwal, I.; Nemoto, K.; Kaiser, C.; Kim, Y.; Malone, T.; "Collective Intelligence in Computer-Mediated Collaboration Emerges in Different Contexts and Cultures;" Bridging Communications; CHI 2015; Seoul Korea
22. Wikipedia Foundation (WF); "Raven's Progressive Matrices;" Oct 2018; https://en.wikipedia.org/wiki/Raven%27s_Progressive_Matrices
23. Wikipedia Foundation (WF); "Wechsler Adult Intelligence Scale;" Oct 2018; https://en.wikipedia.org/wiki/Wechsler_Adult_Intelligence_Scale

24. Wikipedia Foundation (WF); “Intelligence Quotient”; Oct 2018; https://en.wikipedia.org/wiki/Intelligence_quotient
25. Yampolskiy, R.; “Artificial Intelligence Safety and Security;” CRC Press, London/New York; 2019; ISBN: 978-0-8153-6982-0
26. Yampolskiy, R.; “Detecting Qualia in Natural and Artificial Agents;” University of Louisville, 2018
27. Siong, Ch., Brass, M.; Heinze, H.; Haynes, J.; Unconscious Determinants of Free Decisions in the Human Brain; Nature Neuroscience; 13 Apr 2008; http://exploringthemind.com/the-mind/brain-scans-can-reveal-your-decisions-7-seconds-before-you-decideAuthor, F., Author, S., Author, T.: Book title. 2nd edn. Publisher, Location (1999).