Using Tononi Phi to Measure Consciousness of a Cognitive System While Reading and Conversing *

Matthew Iklé[1,3][0000−0001−5978−2703], Ben Goertzel[1,2], Misgana Bayetta[1,2], George Sellman[3], Comfort Cover[3], Jennifer Allgeier[3], Robert Smith[3], Morris Sowards[3], Dylan Schuldberg[4], Man Hin Leung[1,2], Amen Belayneh[1,2], Gina Smith[2], and David Hanson[2]

[1] SingularityNET, Units 3B/4B, Tower 2, South Seas Centre, 75 Mody Road, Kowloon, Hong Kong
http://www.singularitynet.io
[2] Hanson Robotics, Unit 209B, 2/F, Photonics Centre, HKSTP, Pak Shek Kok, N.T., Hong Kong
http://www.hansonrobotics.com
[3] Adams State University, 208 Edgemont Blvd, Alamosa, CO, 81101
http://www.adams.edu
[4] Alamosa High School, 805 Craft DR, Alamosa, CO 81101
http://ahs.alamosa.k12.co.us

Abstract. We conducted computational experiments estimating Tononi's Phi coefficient to measure the "integrated information" within the OpenCog cognitive architecture on two types of tasks: Reading (i.e. parsing and semantically analyzing) short documents, and guiding the Sophia humanoid robot in carrying out a dialogue-based interaction. The data used to calculate Phi comprises time-series of STI (Short Term Importance) values corresponding to Atoms (nodes and links) in OpenCog's Attentional Focus. To make these computations feasible, we preprocessed our data using Independent Component Analysis. We fed the reduced set of time series into software that applies known methods for approximating Phi. Qualitatively (and preliminarily), comparison of the variation of Phi with cognitive system behavior over time reveals sensible patterns.

Keywords: Integrated Information Theory · Tononi Phi · Attention Allocation · System Connectedness · Neural Correlate of Consciousness · Machine Consciousness · Humanoid Robotics

1 Introduction

Measurement and analysis of overall system states over time of complex cognitive AI systems is a significant problem. Analyzing detailed log files rapidly becomes

* Initial research conducted Summer 2018, while Dr. Iklé was Professor of Computer Science and Mathematics at Adams State University. Dr. Iklé oversaw 2 computer science professors, 6 undergraduate interns, and 2 high-school apprentices. The summer program was supported by Army Research Office grant W911NF-15-1-0514.

overwhelming, and it is also necessary to have overall, holistic measurements of system states. Facing this situation with our OpenCog AI system as it reads texts and works within the Hanson AI framework to guide the dialogue of the Sophia humanoid robot 1, we turned to the Tononi Phi coefficient as a tool for measuring the overall level of "integrated information" in the OpenCog system. Phi has been posited by Tononi and others as a fundamental measure of the "level of consciousness" in a system [15]. One of the authors has presented an alternate view in which Phi is one way of approximating one among many aspects of consciousness in certain classes of cognitive systems [5]. Here, it is sufficient to consider Phi as one interesting measure of a consciousness-related property (holistic information integration) of a complex cognitive system.

We conducted two experiments: one with OpenCog parsing and semantically analyzing short documents, and the other with OpenCog controlling the Sophia humanoid robot while leading a person through a structured meditation session. In both experiments we logged the STI values for the Atoms in the system's Attentional Focus, and estimated Phi from the time-series thus obtained.

A core practical tool was Matlab code from Kitazona and Oizumi [7] for estimating Phi from coupled time series. We found this code to be best for analyzing a small number of dense time series, as opposed to the large number of sparse time series presented via exportation of STI time series from the Attentional Focus. Thus we adopted a novel methodology of first applying Independent Component Analysis (ICA) to reduce the original set of sparse time series to a smaller number of dense time series, and then applying the Phi measure. The common foundation of ICA and Phi in the mathematics of mutual information provides some consilience here. We compared the Phi value time-series obtained with the time-series of events in the external situation and behavior of the OpenCog system in the two experiments. Qualitatively we found correspondences between changes in Phi and changes in the situation and behavior of the cognitive system, providing preliminary validation for the methodology pursued.

2 The OpenCog Cognitive Architecture and the Hanson Robotics Sophia Robot

OpenCog is a complex, integrative cognitive AGI architecture currently used in a variety of practical applications, including natural language processing and humanoid robot control [16]. The architecture combines multiple AI paradigms such as uncertain logic, computational linguistics, evolutionary program learning, and connectionist attention allocation in a unified architecture. Cognitive processes embodying these different paradigms interoperate on a common neural-symbolic knowledge store called the Atomspace (a weighted labeled hypergraph whose nodes and links are called "Atoms"). This interaction is designed to encourage the self-organizing emergence of high-level network structures in the Atomspace.

One of OpenCog's core cognitive algorithms is the Economic Attention Networks (ECAN) module for system resource management [2]. ECAN views the

Atomspace as a graph of untyped nodes, and considers links of type HebbianLink, and other links with assumed HebbianLink semantics. Each Atom in ECAN is weighted with two numbers, STI (short-term importance) and LTI (long-term importance.) A system of equations spread importance among atoms based upon the importance of their roles in performing actions related to the system's current goals. An important concept within ECAN is the Attentional Focus, consisting of those Atoms currently deemed most important by the system in terms of achieving its goals. OpenCog has recently been integrated with the Hanson AI framework, that is used to control the Sophia humanoid robot [6]. Sophia's high degree of human likeness coupled with her array of complex emotional expressions make her an ideal platform for cognitive robotics experimentation.

3 Integrated Information Theory and the Tononi Phi Measure

Fig. 1. Image of Sophia and a human conversation partner during a "Loving AI" robot-meditation-guide trial conducted in California in 2018. In this trial, Sophia was running the same OpenCog-based control code as in Experiment 2 reported here.

Created by University of Wisconsin psychiatrist and neuroscientist Giulio Tononi in 2004 [15], Integrated Information Theory (IIT) is an evolving system and calculus for studying and quantifying consciousness. It is strikingly Cartesian [9] in how it approaches the problem: Rather than investigating consciousness by looking at neurons and neurological networks first, as most efforts have [10], it begins by examining lived experience of what it is to be conscious. It then builds an understanding of what consciousness must require neurologically from there.

The IIT that emerges is a detailed, complex framework describing how consciousness behaves and is organized. Its centerpiece is Phi, Tononi's [15] mathematical quantifier for consciousness. Based on the number and quality of interconnections a given entity has between bits of information, Phi is hypothesized by Tononi and colleagues to correspond to how conscious it is [12,14]. Some of the authors incline toward a nuanced view in which Phi is viewed as an estimator of one among many interesting properties of consciousness in roughly human-like cognitive systems [5]. However, practical estimation of Phi in cognitive systems has value regardless of debates over foundational interpretation.

3.1 Procedure for Calculating Phi

In calculating Tononi Phi values, three major issues arise: As Max Tegmark [13] points out, there are at least 420 choices one can make in calculating the measure; Determination of the "Minimum Information Partition" (MIP) of the causal graph structure grows super-exponentially with the number of nodes;

and the size of the probability distribution vectors required to determine Phi also increases super-exponentially with the number of nodes. We have chosen to handle these issues as follows: Two "good" methods for calculating Phi are Φ^*, an approximate measure of Phi, and Phi 3.0, both introduced by Oizumi [13]. We have chosen to implement Phi 3.0, calculating probability distributions according to the procedure described in [8]. Oizumi has empirically demonstrated that Queyranne's Algorithm provides a good approximation of the MIP in $O\left(N^3\right)$ time. We based our Python code upon the Matlab code from Kitazono and Oizumi [7] which integrates Queyranne's Algorithm with the Phi calculation to find the MIP. We stored the (often sparse) probability distributions using Python dictionaries instead of arrays.

In our second experiment we possessed thousands of sparse time series, so we adopted a novel approach, in which we first applied Independent Component Analysis (ICA) to reduce the problem dimensionality; and then calculated Phi from the dimensionally-reduced time series. Since it was unclear how many independent dimensions we should reduce to, we also calculated the sum of the squares of the residuals (SSR) for each dimension, and chose the dimension giving minimum total SSR. This aspect of the methodology will merit from further experimentation and refinement.

4 Measuring OpenCog's Attentional Dynamics During Reading and Dialogue

We now describe two experiments in which we measured Phi values for the STI values obtained from OpenCog's Attentional Focus while carrying out practical tasks.

4.1 Experiment 1: Reading Documents About Insects and Poison

For Experiment 1, we built upon an earlier experiment created to study attentional dynamics during OpenCog-based language comprehension [4]. We first fed the OpenCog Atomspace with prior knowledge about relations between English words, based on the Wordnet and Conceptnet4 databases, and with SimilarityLinks between words with weights calculated using the Adagram neural network [3]. We started the ECAN system and the NLP pipeline, and loaded articles into the Atomspace, beginning with articles regarding insects and shifting to articles regarding poisons. As the system ingests each sentence, WordNodes corresponding to each word are stimulated with STI, thus triggering attentional focus dynamics correlated with the reading process. One goal of the study was to observe whether, after reading documents regarding insects and then poisons, attention would spread to concepts related to insecticide. This phenomenon did occur, as shown in Figure 2.

We also calculated Phi values based upon the ConceptNodes "insect", "poison", and "insecticide." As Figure 3 shows, there was an interesting jump in the Phi value when "insecticide" first became important, suggesting that the Phi increase

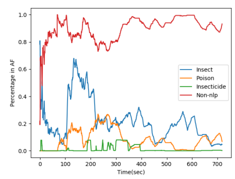

Fig. 2. Varying prevalence in the Attentional Focus of Atoms related to insects, poison and insecticide while OpenCog reads a series of short documents focused on insects and then poison.

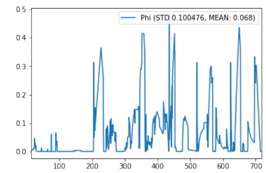

Fig. 3. Phi estimated from the Attentional Focus while OpenCog reads a series of short articles pertinent to insects and poison.

was correlated with an increased complexity of attentional spreading within the Atomspace.

4.2 Experiment 2: Meditation-Guiding Dialogue

In Experiment 2 we used OpenCog and the Ghost dialogue-control framework [1] of the Hanson AI system to control the Sophia robot. The Ghost script enabled Sophia to conduct part of a guided meditation session [11]. We found thousands of Atoms passing through the Attentional Focus, and thus leveraged the ICA-based methodology described above, using an optimal embedding dimension of 3. The Phi time series obtained are shown in Figure 4. Phi is higher while verbal interaction is occurring and lower while Sophia is watching her subject meditate or breathe deeply, etc.

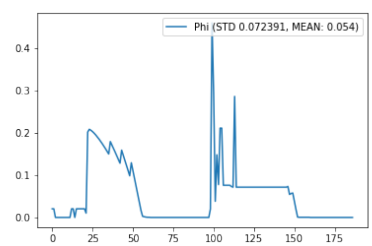

Fig. 4. Phi estimated from OpenCog while controlling Sophia conducting dialogue to guide a meditation session.

5 Conclusion and Future Work

Our experiments appear to demonstrate some connection between sensible behavior and higher Phi values, providing preliminary validation of our methodology. The next step is to conduct significantly more extensive experimentation, as well as to further elaborate the theoretical properties of the "ICA plus Phi" pipeline. We are also interested in improving our software to enable real-time calculation of Phi during system operation, and using Phi as a target for adaptive optimization of ECAN parameters.

References

1. (November 2018), https://github.com/opencog/opencog/tree/master//opencog/ghost
2. B., G., Hitzler, P., Hutter, M. (eds.): Economic Attention Networks: Associative Memory and Resource Allocation for General Intelligence, vol. 171. IOS Press (2008). https://doi.org/10.2991/agi.2009.19
3. Bartunov, S., Kondrashkin, D., Osokin, A., Vetrov, D.P.: Breaking sticks and ambiguities with adaptive skip-gram. CoRR **abs/1502.07257** (2015), http://arxiv.org/abs/1502.07257
4. Belachew, M., Goertzel, G., Iklé, M., Hanson, D.: Measuring integrated information from the decoding perspective. Biologically Inspired Cognitive Architectures **25**, 130–134 (August 2018). https://doi.org/10.1016/j.bica.2018.07.005
5. Goertzel, B.: Characterizing human-like consciousness: An integrative approach. Procedia Computer Science **41**, 152–157 (2014). https://doi.org/10.1016/j.procs.2014.11.098
6. Goertzel, B., Mossbridge, J., Monroe, E., Hanson, D., Yu, G.: Humanoid robots as agents of human consciousness expansion (2017), https://arxiv.org/pdf/1709.07791
7. Kitazono, J., Oizumi, M.: Practical PHI Toolbox (9 2018). https://doi.org/10.6084/m9.figshare.3203326.v10, https://figshare.com/articles/phi_toolbox_zip/3203326
8. Krohn, S., Ostwald, D.: Computing integrated information. Neuroscience of Consciousness **2017**(1), nix017 (2017). https://doi.org/10.1093/nc/nix017, http://dx.doi.org/10.1093/nc/nix017
9. Leffler, T.: The hard problem of consciousness: A mathematical approach. arXiv:1704.01148v2 [q-bio.NC] (2017)
10. Massimini, M., Tononi, G.: Sizing up Consciousness: Towards an objective measure of the capacity for experience. Oxford University Press (2018)
11. Mossbridge, J., Goertzel, B., Mayet, R., Monroe, E., Nehat, G., Hanson, D., Yu, G.: Emotionally-sensitive ai-driven android interactions improve social welfare through helping people access self-transcendent states. vol. AI for Social Good Workshop at Neural Information Processing Systems 2018 Conference (2018)
12. P., M.W.G., Marshall, W., Albantakis, L., Findlay, G., Marchman, R., Tononi, G.: Pyphi: A toolbox for integrated information theory. PLoS Comput Biol (2018). https://doi.org/10.1371/journal.pcbi.1006343
13. Tegmark, M.: Improved measures of integrated information. PLOS Computational Biology **12**(11), 1–34 (11 2016). https://doi.org/10.1371/journal.pcbi.1005123, https://doi.org/10.1371/journal.pcbi.1005123
14. Tononi, G.: Consciousness as integrated information: a provisional manifesto. The Biological Bulletin **215**(3), 216–242 (December 2008)
15. Tononi, G.: An information integration theory of consciousness. BMC Neuroscience **5**(1), 42 (2004). https://doi.org/10.1186/1471-2202-5-42, http://www.biomedcentral.com/1471-2202/5/42
16. Wang, P., Goertzel, B., Franklin, S. (eds.): OpenCog: A Software Framework for Integrative Artificial General Intelligence, vol. 171. IOS Press (2008)