Informational Theories of Consciousness: A Review and Extension

Igor Aleksander[1] and David Gamez[2]

[1] Department of Electrical Engineering, Imperial College, London SW7 2BT, UK
i.aleksander@imperial.ac.uk

[2] Department of Electrical Engineering, Imperial College, London SW7 2BT, UK
david@davidgamez.eu

Abstract In recent years a number of people have suggested that there is a close link between conscious experience and the differentiation and integration of information in certain areas of the brain. The balance between differentiation and integration is often called information integration, and a number of algorithms for measuring it have been put forward, which can be used to make predictions about consciousness and to understand the relationships between neurons in a network. One of the key problems with the current information integration measures is that they take a lot of computer processing power, which limits their application to networks of around a dozen neurons. There are also more general issues about whether the current algorithms accurately reflect the consciousness associated with a system. This paper addresses these issues by exploring a new automata-based algorithm for the calculation of information integration. To benchmark different approaches we have implemented the Balduzzi and Tononi algorithm as a plugin to the SpikeStream neural simulator, which is used in this paper to carry out some preliminary comparisons of the liveliness and Φ measures on a simple four neuron network.

1. Introduction

In recent years a number of people have suggested that there is a close link between conscious experience and the balance between the differentiation and integration of information in certain areas of the brain. This combination of integration and differentiation is often called information integration, and Tononi has claimed that: "at the fundamental level, consciousness is integrated information, and … its quality is given by the informational relationships generated by a complex of elements" (Tononi 2008, p. 217). A number of algorithms for measuring information integration have been put forward, which can be used to make predictions about consciousness and to assist with the debugging of artificial cognitive systems (Gamez and Aleksander 2009).

A better understanding of what information integration actually is can be gained by considering an example of a digital camera sensor with a million photodiodes (Tononi 2008). If each photodiode is binary, the sensor can enter a total of $2^{1,000,000}$ states, which corresponds to 1,000,000 bits of information. One of the key differences between the photodiode and the areas of the brain associated with consciousness is that each photodiode acts independently of the others, whereas the vast number of states in the brain's neural networks are the outcome of causal interactions between the neurons. Both the brain and the camera sensor can enter a large number of states, but Tononi (2008) claims that some of the brain's states are conscious because they are both differentiated *and* integrated at the same time. This combination of differentiation and integration is illustrated in Figure 1, which contrasts conscious systems, such as parts of the brain, with systems that lack integration between their elements, or which can only enter a low number of states.

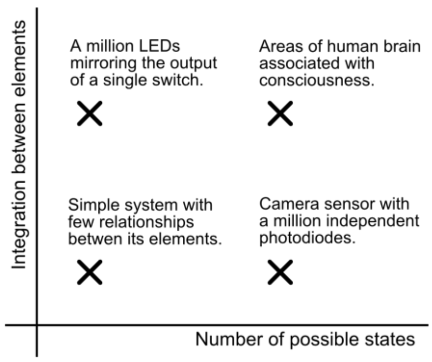

Fig. 1. Systems with different amounts of integration and differentiation. At the bottom left are systems with few states and low integration between these states, such as simple creatures or basic artificial systems. Bottom right are highly differentiated systems with little or no integration, such as the photodiodes in a digital camera sensor. Top left are systems whose high level of integration prevents them from entering a wide variety of states – for example, a large number of LEDs reflecting the state of a single switch. Top right are systems with a large repertoire of states that are the result of causal interactions between the elements, such as the areas associated with consciousness in the human brain.

A number of algorithms have been put forward to measure information integration (Tononi and Sporns, 2003; Balduzzi and Tononi, 2008), and there are a number of related measures including neural complexity (Tononi et al. 1994, 1998), transfer entropy (Schreiber, 2000) and causal density (Seth et. al., 2006). A major problem with many of these algorithms is that they can take an extremely long time to compute. For example, the algorithms put forward by Tononi and Sporns (2003) and Balduzzi and Tononi (2008) have multiple factorial dependencies, which require calculations on all possible partitions of all possible subsets of a network. It has been estimated that a full analysis of an 18,000 network using Tononi and Sporns' (2003) algorithm would take 10^{9000} years (Gamez 2008), and the measured and predicted times for the more recent Balduzzi and Tononi (2008) algorithm are also extremely large (see Figure 3). Much more efficient ways of calculating information integration need to be found if it is to become a useful tool in research on consciousness and robotics.

A second key issue is that different information integration algorithms are likely to make different predictions about which parts of a system are conscious, and more work is needed to determine the accuracy of the current measures. This issue will be easier to address when the performance issues have been solved because the current algorithms are typically only capable of analyzing systems with around a dozen elements. More discussion is also needed about Tononi's strong claim that consciousness *is* information integration, which will be easier to address if it can be empirically established whether information integration is consistently linked with conscious states.

To address these issues we are exploring an alternative way of calculating information integration based on liveliness and an automata approach (Aleksander and Atlas, 1973), which is outlined in Section 4. To benchmark different information integration algorithms we have implemented the Balduzzi and Tononi (2008) algorithm as a plugin to the SpikeStream neural simulator (see Section 3), and Section 5 describes some very preliminary experiments in which a simple 4 neuron network was analyzed using the two different algorithms.

2. Previous Work

Other analysis work based on information integration has been carried out by Lee et al. (2009), who made multi-channel EEG recordings from eight sites in conscious and unconscious subjects and constructed a covariance matrix of the recordings on each frequency band that was used to identify the complexes within the 8 node network using Tononi and Sporns' (2003) algorithm. This experiment found that the information integration capacity of the network in the gamma band was significantly higher when subjects were conscious. Information integration-based predictions about the consciousness of an 18,000 neuron network have been carried out by Gamez (2010) and there has been some theoretical work on information integration by Seth et al (2006), who identified a number of weaknesses in Tononi and Sporns' (2003) method and criticized the link between information integration and consciousness.

A number of other measures of the information relationships between neurons have been put forward, including neural complexity (Tononi et al. 1994, 1998), transfer entropy (Schreiber, 2000) and causal density (Seth et. al., 2006). There has been some work comparing neural complexity measures and graph theory (Shanahan, 2008), and these measures have been used by a number of people to examine the anatomical, functional and effective connectivity of biological networks, either using scanning or electrode data, or large-scale models of the brain. One example of this type of work is Honey et al. (2007), who used transfer entropy to study the relationship between anatomical and functional connections on a large-scale model of the macaque cortex, and demonstrated that the functional and anatomical connectivity of their model coincided on long time scales. Another example is Brovelli et al. (2004), who used Granger causality to identify the functional relationships between recordings made from different sites in two monkeys as they pressed a hand lever during the wait discrimination task, and Friston et al. (2003) modeled the interactions between different brain areas and made predictions about the coupling between them. Information-based analyses have also been used to guide and study the evolution of artificial neural networks connected to simulated robots (Seth and Edelman, 2004; Sporns and Lungarella, 2006). An overview of this type of research can be found in Sporns et. al. (2004) and Sporns (2007).

3. SpikeStream Analysis Software

To evaluate different approaches to the calculation of information integration we have developed a plugin analysis framework for the SpikeStream simulator (Gamez 2007). SpikeStream was originally developed as an open source spiking neural simulator, and we have adapted it to support weightless neurons and added a plugin to carry out the analysis of networks using Balduzzi and Tononi's (2008) information integration algorithm. To save development time we temporarily removed the simulation functions of SpikeStream, and have been using the NRM weightless neural simulator[1] to build and run networks that can be imported into SpikeStream for analysis. This framework makes it reasonably easy to add other plugins to carry out different types of analysis. A screen shot of the plugin implementing Balduzzi and Tononi's (2008) algorithm is shown in Figure 2.

[1] More information about NRM is available at: http://www.iis.ee.ic.ac.uk/eagle /barry_dunmall.htm.

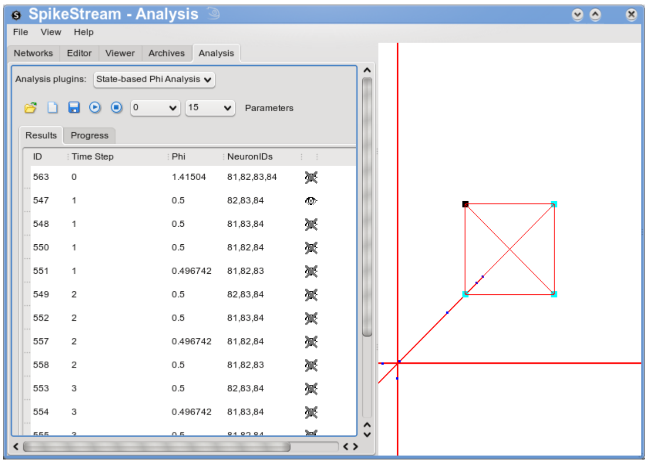

Fig. 2. SpikeStream analysis plugin implementing Balduzzi and Tononi's (2008) algorithm. The graphical controls are on the left above a table listing the analysis results. The 3D display on the right shows the network that has been analyzed, with one of its complexes highlighted.

To evaluate the performance of the Balduzzi and Tononi (2008) algorithm the analysis times for different network sizes were recorded and plotted in the graph shown in Figure 3. These were extrapolated to generate predicted times for the analysis of larger networks.

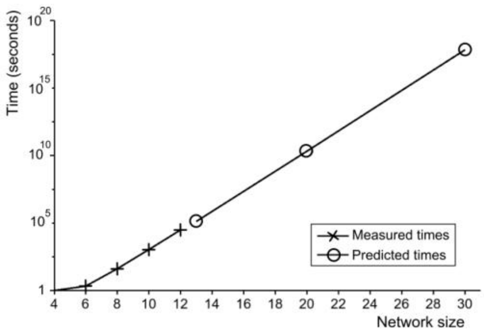

Fig. 3. Measured and predicted times for the calculation of information integration on different sizes of network using Balduzzi and Tononi's (2008) algorithm and a Pentium IV 3.2 GHz single core computer. Each neuron in the network was randomly connected to five other neurons and their truth tables had five entries. The results are for the analysis of a single time step with a random firing pattern.

The graph in Figure 3 shows a very rapid increase in the time taken for the analysis, with networks of 30 neurons being estimated to take 10^{10} years to analyze on a desktop computer. Whilst some optimizations might be able to increase the calculable network size, this performance issue cannot be entirely overcome because of the factorial dependencies of the calculation. This problem is corroborated by earlier work (Gamez 2008), which estimated that it would take 10^{9000} years to analyze a network with 18,000 neurons using Tononi and Sporns (2003) algorithm.

We are currently preparing a release of the SpikeStream analysis software, which will be available in Windows and Linux versions from the SpikeStream website (http://spikestream.sf.net). The next step will be to explore more efficient ways of calculating information integration, which can be benchmarked and compared by implementing them as SpikeStream plugins.

4. Some Finite-State Discrete Automata Principles

Balduzzi and Tononi's (2008) recent work on discrete-state systems have made it possible to examine information integration ideas using classical automata theory. This section outlines some principles of the automata approach, which is compared with Balduzzi and Tononi's (2008) algorithm in Section 5.

4.1 Cyclic Activity and Neuron Liveliness

In 1973, Aleksander and Atlas carried out an analysis of epigenesis and differentiation in networks of interconnected genes, which were modeled by 2-input binary neurons with fixed, but randomly chosen, neuron transfer functions. When these networks were started in a random state, they went from state to state until they returned to a previously visited state, at which point the cycle was complete and repeated itself in perpetuity. In this model the states in the cycle were analogous to the chemical changes in the division of a cell, with the number of different cycles involved in a given network being analogous to the number of different cell types of a particular organism. Aleksander and Atlas (1973) showed how this type of network was stable, even if its elements were interconnected at random.

In Aleksander and Atlas (1973) the concept of *liveliness* was defined for two-input network elements or 'nodes'. In this paper the concept of liveliness will be presented in terms of an arbitrary number of *n* inputs. Let the binary inputs of a node *j* be: $x_1^j, x_2^j \ldots x_n^j$. The vectors of the 1 or 0 states of these inputs can be represented as $X_0^j, X_1^j \ldots X_{2^n-1}^j$, where, for example, if n=3 and k=5, $X_k^j = \overline{x}_1^j x_2^j \overline{x}_3^j$ signifies that the inputs are 010. A node is said to be *lively* on input x_p^j if its output z^j changes value when x_p^j changes its value. This is denoted by $\lambda(x_p^j) = 1$, or $\lambda(x_p^j) = 0$ if it is not lively. Liveliness is a function of the choice of X_k^j and it can be computed for all of the 2^n possible vectors.

The *liveliness of a node* for an input x_p^j, $\Lambda(x_p^j)$, is the number of times for all choices of *k* in X_k^j that the node is lively divided by the number of input combinations, 2^n. As an example consider the function described in Table 1.

x1	x1	x1	z	$\lambda(x_1^j)$	$\lambda(x_2^j)$	$\lambda(x_3^j)$
0	0	0	0	1	0	1
0	0	1	1	0	1	0
0	1	0	0	0	0	0
0	1	1	0	0	1	0
1	0	0	1	1	1	0
1	0	1	1	0	1	0
1	1	0	0	0	1	0
1	1	1	0	0	1	0

Table 1. Example function to illustrate the calculation of liveliness

The total node liveliness for the inputs in Table 1 is as follows:

$$\Lambda(x_1^j) = 2/8; \qquad \Lambda(x_2^j) = 6/8; \qquad \Lambda(x_3^j) = 1/8.$$

Given a group of nodes the liveliness of a loop, *LΛ*, can be found by taking the product of the liveliness of the connections in a path traversing the nodes. The liveliness of a node, *NΛ*, is defined as the average liveliness of all the inputs to a node (scoring 0 for disconnections). Nodes that compute the parity function (output 1 if the number of 1s at the input is even) or the 'oddness' function (output 1 if the number of 1s at the input is odd) have the maximum liveliness of 1 on all inputs, while 'stuck at 1' or 'stuck at 0' functions have a liveliness of 0. Common functions such as AND and OR can be shown to have a liveliness of $2/2^n$.

In Aleksander and Atlas (1973) this approach was used to predict the probability of finding closed element loops, whose ability to transmit information caused the cyclic state structures from which the cyclic activity of genetic networks could be predicted. This was seen to be a close approximation to what is known about epigenesis and differentiation across animal species.

4.2 Relevance to information integration.

While a formal link between liveliness and information integration is work in progress, it is possible to identify a heuristic relationship between them. The high liveliness between nodes that are in closer causal contact appears to broadly correspond to the causal interdependence between neurons that is measured by Φ. For example, the digital camera sensor mentioned earlier exhibited low Φ because there was no causal linking between the elements, and the overall liveliness of this system would be zero as well. However, whilst Φ and liveliness both measure causal interactions, they do not measure exactly the same thing: Φ indicates what information a system provides when it enters a particular state, whereas liveliness identifies the state structures that arise as a result of the causal interactions available to the neurons.

5. Examples

The experiments in this section were carried out on a 4-neuron network (A,B,C,D) that was constructed with different functions and connectivity. The SpikeStream analysis software was used to identify the networks' complexes and measure their Φ, and these results were compared to the liveliness of each system.

5.1 Example 1

The connections in the first network are shown in Fig. 4A, which had an AND function in neuron C. The liveliness of this network's connections are shown in Fig. 4B.

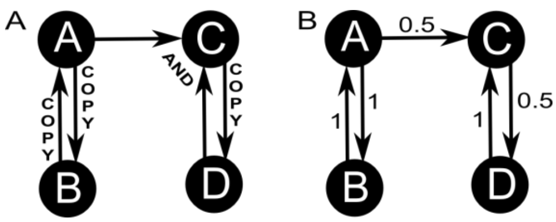

Fig. 4. (A) Connections and functions of Network 1; (B) Liveliness of connections in Network 1

The Φ analysis of Network 1 was carried out for each of the 16 possible states of the network. The highest Φ complex of each set of connected neurons was identified for each state of the network, and these were averaged to produce the results shown in Table 2.

Complex	AB	CD	AC	ACD	ABCD
Average Φ (bits)	2	1.5	1.0	0.7	0.3
Loop Λ (%)	100	50	50	25	25

Table 2. Φ and Loop Λ values for Network 1

These results show a rough correlation between the loop liveliness of a set of connected elements and the average highest Φ value of the complexes in which they are involved. This suggests the possibility that the complexes discovered by the exhaustive processor-intensive calculations could be approximated using the liveliness approach.

5.2 Example 2

In Network 2 the AND function in neuron C in Network 1 was replaced by an XOR function (see Fig. 5A), which led to the loop and node liveliness values shown in Fig. 5B.

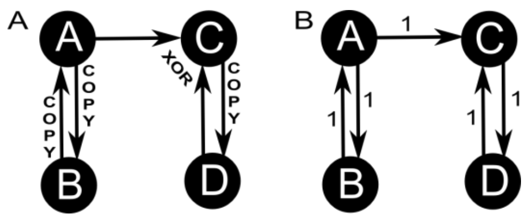

Fig. 5. (A) Functions and connections of Network 2; (B) Liveliness of connections in Network 2

Complex	AB	ACD	ABCD
Average Φ (bits)	2	2	1
Loop Λ (%)	100	100	100
Node Λ (%)	11	7	2.2

Table 3. Φ, loop Λ, and node Λ values for Network 2

When the results in Table 3 are compared with the results in Table 2 it can be seen that the higher loop liveliness caused by the XOR function is roughly mirrored in higher Φ values of the complexes AB, ACD and ABCD. The results also show that average node liveliness is a better fit with the Φ data, which could be investigated more systematically in future work.

5.3 Example 3

This experiment took the highest Φ values of the complex ABCD for different network configurations and compared it to the average node liveliness. The results are shown in Table 4.

Connections	Functions	Max Φ (bits)	Average Node Λ (%)
Same as Fig. 4.	AND function between A and C	0.58	25
Same as Fig. 4	XOR function between A and C	1.0	31
Fig. 4 network with two way connections between AB, AC, BD, and DC.	All nodes with XOR function	0.83	50
Two way connections between each node and all other nodes.	All nodes with XOR function	4	100

Table 4. Comparison of Φ and average node Λ values for ABCD with different connections and functions

In these results the maximum value of Φ was associated with the maximum liveliness (row 4) and the minimum value of Φ was associated with the minimum liveliness (row 1). However, the values in rows 2 and 3 are not correlated and more work needs to be done to establish the extent to which liveliness and Φ differ and which is the best measure of information integration in a system.

It must be stressed that this work on the link between Φ and Λ is highly speculative at this stage, and the examples are only intended as a highly embryonic heuristic, which might eventually lead to better ways of measuring information integration.

6. Future Work

One of the first things that is needed to take this work forward is a set of networks that can be used to benchmark and compare different methods of calculating information integration. These networks would need to be of increasing size and complexity and their topologies should be likely to exhibit different levels of information integration. These networks could be used to measure the time performance of each algorithm, and the high integration areas could be compared. Once a sensible set of benchmarks has been defined, it will be possible to develop more efficient information integration algorithms that accurately reflect the relationships between elements in a network.

Improved measures of information integration have many applications. Within neurophenomenology, information integration can be used to make predictions about what a person is conscious of based on fMRI or other data, and it is also possible to use information integration to make predictions about the consciousness of artificial systems as part of work on synthetic phenomenology (Gamez, 2010). Information integration can also be used to develop representations of the contents of the minds of artificial systems, which will be very useful for debugging cognitive robots that learn through interaction with their experiences (Gamez and Aleksander, 2009).

7. Conclusions

The measurement of information integration is becoming increasingly important in the study of natural and artificial consciousness, and it also has applications in network analysis and the debugging of cognitive systems. The two key challenges in this area are the development of faster information integration algorithms and the validation of these algorithms on non-trivial systems. This paper has outlined a new approach to the measurement of information integration based on liveliness and described software with a plugin framework that can be used to benchmark different types of information integration algorithm. The second half of this paper offered some preliminary comparisons between the Φ and liveliness of a simple four neuron network. These examples are very tentative and would need to be developed in much more detail to accurately evaluate the advantages and disadvantages of the liveliness approach.

Acknowledgments This work was supported by a grant from the *Association for Information Technology Trust*.

References

Aleksander I, Atlas P (1973) Cyclic Activity in Nature: Causes of Stability. International Journal of Neuroscience 6: 45-50.

Balduzzi D, Tononi G (2008) Integrated Information in Discrete Dynamical Systems: Motivation and Theoretical Framework. PLOS Computational Biology 4(6): e1000091.

Brovelli A, Ding M, Ledberg A, Chen Y, Nakamura R, Bressler SL (2004) Beta oscillations in a large-scale sensorimotor cortical network: Directional influences revealed by Granger causality. Proc Natl Acad Sci USA 101: 9849-54.

Friston KJ, Harrison L, Penny W (2003) Dynamic causal modelling. NeuroImage 19: 1273–302.

Gamez D (2007) SpikeStream: A Fast and Flexible Simulator of Spiking Neural Networks. In J. Marques de Sá, L.A. Alexandre, W. Duch and D.P. Mandic (eds.), Proceedings of ICANN 2007, Lecture Notes in Computer Science Volume 4668, Springer Verlag, pp. 370-9.

Gamez D (2008) The Development and Analysis of Conscious Machines. Unpublished PhD thesis, University of Essex, UK. Available at: http://www.davidgamez.eu/mc-thesis/.

Gamez D (2010) Information Integration Based Predictions about the Conscious States of a Spiking Neural Network. Consciousness and Cognition, doi:10.1016/j.concog.2009.11.001.

Gamez D, Aleksander I (2009) Taking a Mental Stance Towards Artificial Systems. Biologically Inspired Cognitive Architectures. Papers from the AAAI Fall Symposium. AAAI Technical Report FS-09-01: 56-61.

Honey CJ, Kötter R, Breakspear M, Sporns O (2007) Network structure of cerebral cortex shapes functional connectivity on multiple time scales. PNAS 104(24): 10240–5.

Lee U, Mashour GA, Kim S, Noh G-J, Choi B-M (2009) Propofol induction reduces the capacity for neural information integration: Implications for the mechanism of consciousness and general anesthesia. Consciousness and Cognition 18(1): 56-64.

Schreiber T (2000) Measuring Information Transfer. Physical Review Letters 85(2): 461-4.

Seth AK, Edelman, GM (2004) Environment and behavior influence the complexity of evolved neural networks. Adaptive Behavior 12: 5-20.

Seth AK, Izhikevich E, Reeke GN, Edelman GM (2006) Theories and measures of consciousness: An extended framework. PNAS 103(28): 10799–804.

Shanahan M (2008) Dynamical complexity in small-world networks of spiking neurons. Phys. Rev. E 78, 041924.

Sporns O (2007) Brain Connectivity. Scholarpedia 2(10):4695.

Sporns O, Lungarella M (2006) Evolving coordinated behavior by maximizing information structure. In L. Rocha, L. Yaeger, M. Bedau, D. Floreano, R. L. Goldstone and A. Vespignia-ni (eds.), Artificial Life X: Proceedings of the 10th International Conference on the Simulation and Synthesis of Living Systems, Cambridge, MA: MIT Press, pp. 322-329.

Sporns O, Chialvo DR, Kaiser M, Hilgetag CC (2004) Organization, development and function of complex brain networks. TRENDS in Cognitive Sciences 8(9): 418-25.

Tononi G (2008) Consciousness and Integrated Information: a Provisional Manifesto. Biol. Bull. 215: 216–242.

Tononi G and Sporns O (2003) Measuring information integration. BMC Neuroscience 4:31.

Tononi G, Sporns O, Edelman GM (1994) A measure for brain complexity: Relating functional segregation and integration in the nervous system. Proc. Natl. Acad. Sci. USA 91: 5033-7.

Tononi G, Edelman, GM, Sporns O (1998) Complexity and coherency: integrating information in the brain. Trends in Cognitive Sciences 2(12): 474-84.