Causal networks in neural systems: From water mazes to consciousness

Anil K Seth
The Neurosciences Institute
10640 John Jay Hopkins Drive
San Diego, CA 92121, USA
Email: anil.k.seth@gmail.com

***Abstract*— Neurons engage in causal interactions with one another and with the surrounding body and environment. Neural systems can therefore be analyzed in terms of causal networks, without assumptions about information processing, neural coding, and the like. Here, we describe the analysis of causal networks in simulated neural systems using a combination of time-series analysis ("Granger causality") and network theory. Implications are drawn for causal pathways in the hippocampus, for the relation between synaptic plasticity and behavioral learning, and for the neural dynamics underlying consciousness.**

I. Introduction

A basic fact about neural systems is that their elements enter into causal interactions with one another, as well as with the surrounding body and environment. Because neural systems are generally composed of large numbers of elements, a useful analysis of their causal interactions must involve *causal networks*. Neural systems can be analyzed in terms of causal networks without assumptions about whether or not they are 'information processing' devices [1], or whether or not there exist 'neural codes' [2].

In this paper, time-series analysis techniques based on "Granger causality" [3] are combined with network theory in order to characterize causal networks in simulated neural systems [4], [5], [6], [7]. We analyze neural simulations, as opposed to empirical data [8], [9], in order to best expose the utility of a causal network perspective. The results provide heuristics for interpreting empirical data and raise a number of specific hypotheses.

After reviewing the principles of Granger causality analysis [3], [4], we describe its application to a complex embodied neural simulation - a *brain based device* (BBD) [10] - that incorporates a detailed model of the mammalian hippocampus [11], [5]. The BBD learns a task similar to a classical experimental paradigm in which rodents are trained to locate a 'hidden platform' in a pool of milky water (a Morris 'water maze' [12]); a task for which an intact hippocampus is required. The analysis reveals causal pathways involving the hippocampus that mediate sensory input and motor output; it also suggests how these pathways are modulated as a result of learning.

We then show how an extension of the above analysis can be used to distinguish causal networks in complex neural populations that lead to specific outputs [6]. The concept of a *causal core* is introduced to refer to the set of neuronal interactions that are significant for a given output, as assessed by Granger causality. Illustrated by application to the same BBD, this analysis reveals that large repertoires of neural interactions contain comparatively small causal cores and that these causal cores become smaller during learning, a result which may reflect the selection of specific causal pathways from diverse repertoires.

Finally, we discuss the application of a causal network perspective to the neural dynamics underlying consciousness [7]. At least in humans, and very likely in other mammals as well [13], some causal interactions in the brain contribute to conscious scenes, whereas others do not. We analyze the suggestion that a particular measure applicable to causal networks, *causal density*, may provide a useful means of quantifying the complexity of the neural dynamics relevant to consciousness [7]. A system with high causal density is one that exhibits a dynamical balance between integration and differentiation, just as conscious scenes are themselves both differentiated (each is composed of many parts and is therefore unique) and integrated (each is experienced as a unified whole) [14], [15], [16].

Taken together, the research reviewed in this article indicates that a causal network perspective provides both novel concepts for understanding neural systems, and a set of practically applicable methods for illustrating, refining, and testing the usefulness of these concepts.

II. Granger causality

The concept of Granger causality is based on prediction: If a signal X_1 causes a signal X_2, then past values of X_1 should contain information that helps predict X_2 above and beyond the information contained in past values of X_2 alone [3]. In practice, Granger causality can be tested using multivariate regression modelling. For example, suppose that the temporal dynamics of two time series, $X_1(t)$ and $X_2(t)$ (both of length T), can be described by a bivariate autoregressive model:

$$X_1(t) = \sum_{j=1}^{p} A_{11,j} X_1(t-j) + \sum_{j=1}^{p} A_{12,j} X_2(t-j) + E_1(t)$$

$$X_2(t) = \sum_{j=1}^{p} A_{21,j} X_1(t-j) + \sum_{j=1}^{p} A_{22,j} X_2(t-j) + E_2(t)$$

where p is the maximum number of lagged observations included in the model (the model order, $p < T$), A contains the coefficients of the model (i.e., the contributions of each lagged observation to the predicted values of $X_1(t)$ and $X_2(t)$), and E_1, E_2 are the residuals (prediction errors) for each time series. If the variance of E_1 (or E_2) is reduced by the inclusion of the X_2 (or X_1) terms in the first (or second) equation, then it is said that X_2 (or X_1) *Granger-causes* X_1 (or X_2). In other words, X_2 Granger-causes X_1 if the coefficients in A_{12} are jointly significantly different from zero. This can be tested by performing an F-test of the null hypothesis that $A_{12} = 0$, given assumptions of covariance stationarity on X_1 and X_2. The magnitude of a Granger causality interaction can be estimated by the logarithm of the corresponding F-statistic [17].

Importantly, the concept of Granger causality can be extended to the n variable case (where $n > 2$), by estimating an n variable autoregressive model. In this case, X_2 Granger-causes X_1 if knowing X_2 reduces the variance in X_1's prediction error when the activities of all other variables $X_3 \ldots X_n$ are also taken into account.

Significant Granger causality interactions between variables can be represented as edges in a graph, allowing the application of graph-theoretic techniques [4]. Because Granger causality is in general not symmetric, these edges will be directed. As shown below, the resulting graphs, or causal networks, can provide an intuitive and valuable representation of functional connectivity within a system.

III. Causal networks in a brain-based device

A useful illustration of causal network analysis is given by its application to Darwin X [11], [5]. Darwin X is a brain-based device (BBD), that is, a physical device which interacts with a real environment via sensors and motors, whose behavior is guided by a simulated nervous system incorporating aspects of the neuroanatomy and neurophysiology of the mammalian hippocampus and surrounding areas (see Fig. 1). Full details of the construction and performance of Darwin X are given in [11], [5]; for present purposes it is sufficient to mention only the following. Darwin X contained analogs of several mammalian brain areas including subareas of the hippocampus and three sensory input streams which receive input from a CCD camera and from odometry (see Fig. 1). Each neuronal unit in Darwin X was taken to represent a group of ≈100 real neurons and was simulated using a mean firing rate model. Synaptic plasticity was implemented using a modified version of the BCM learning rule [18] in which synapses between strongly correlated neuronal units are potentiated and synapses between weakly correlated neuronal units are depressed. In some pathways, synaptic changes are further modulated by the activity of a simulated *value system* (area S in Fig. 1) which responded to salient events (see below). The full Darwin X model contained 50 neural areas, ≈90,000 neuronal units, and ≈1,400,000 synaptic connections.

Darwin X was trained on a 'dry' version of the Morris water maze task [12], in which the device learned to locate a 'hidden platform' in a rectangular arena with diverse visual landmarks hung on the walls. Darwin X could only detect the hidden platform when it was directly overhead, by means of a downward facing infrared sensor. Each encounter with the platform stimulated the value system which modulated synaptic plasticity in the value-dependent pathways of Darwin X's simulated nervous system. Darwin X was trained over 17 'trials', each beginning from one of four initial positions. Initially, Darwin X moved randomly, but after about 10 trials, the device reliably took a comparatively direct path to the platform from any starting point [11], [5].

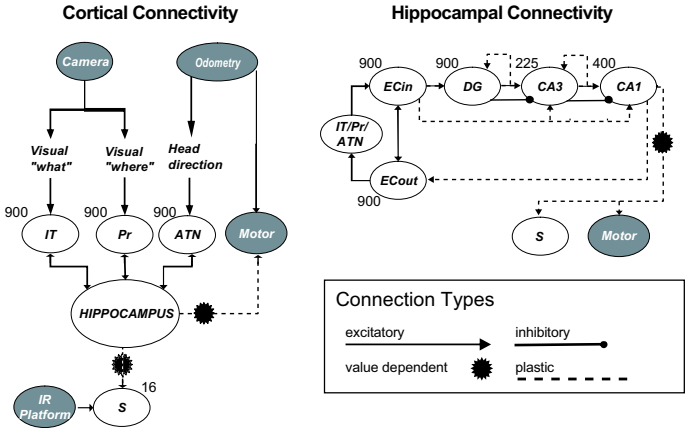

Fig. 1. Schematic of Darwin X's simulated nervous system. There were two visual input streams responding to the color (IT), and width (Pr), of visual landmarks on the walls, as well as one odometric input signalling Darwin X's heading (ATN). These inputs were reciprocally connected with the hippocampus which included 'entorhinal' cortical areas EC_{in} and EC_{out}, 'dentate gyrus' DG, and the $CA3$ and $CA1$ hippocampal subfields. The number of simulated neuronal units in each area is indicated adjacent to each area. This figure is reprinted from [6].

Causal interactions in Darwin X were analyzed by selecting 'functional circuits' as follows (see [5] for details). For a given neuronal unit in $CA1$ (the *reference unit*), a set of neuronal units was selected by identifying those units (from different neural areas) that covaried the most in their activity with the reference unit. The activity time-series corresponding to these functional circuits were then analyzed using a multivariate Granger causality analysis. Fig. 2 shows patterns of causal interactions from a representative reference unit both early in learning (left) and late in learning (right). After learning, the causal network involving the reference unit is much denser and has developed a so-called *trisynaptic loop* in which sensory signals from cortex follow a chain of causal influences through the various subareas of the hippocampus before returning to cortex.

Repeating the above analysis 400 times (see [5]), using each $CA1$ neuronal unit as a reference unit, revealed (i) an increase in the proportion of causal 'shortcuts'; in these shortcuts (which reflect the so-called *perforant pathway*) signals from cortex causally influenced the $CA1$ reference unit without causally involving the intermediate stages of DG and $CA3$, and (ii) after learning, the causal influence exerted by neuronal units in ATN (reflecting head direction signals) grew

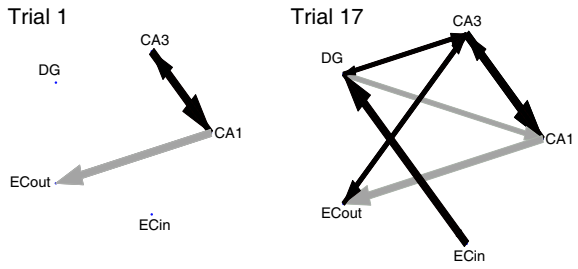

Fig. 2. Causal connectivity patterns for a representative $CA1$ reference unit during the first trial (left) and the last trial (right). Grey arrows show unidirectional connections and black arrows show bidirectional connections. The width of each arrow (and size of arrowhead) reflect the magnitude of the causal interaction. This figure is adapted from [5].

markedly. These observations together suggest that as learning progressed, Darwin X relied less on integrating multisensory signals and its behavior was increasingly driven by odometric signals with modulation from visual areas. This hypothesis may be testable in future animal experiments.

IV. 'Causal Cores' in Neural Populations

The causal networks described above were identified using a multivariate analysis of a small number of neuronal units from different regions of Darwin X's simulated nervous system. We turn now to a variant of this analysis which distinguishes causal interactions within large neural populations that lead to specific outputs [6]. Because this analysis requires complete knowledge of neuroanatomy and dynamics of the studied system, it is also well illustrated by application to Darwin X.[1]

The general framework for this analysis is illustrated in Fig. 3. First, a *neural reference* (NR) is selected from among many possible neuronal events, in this example by virtue of its relationship to a specific behavioral output (Fig. 3A). In the terminology introduced above, a NR refers to the activity of a reference unit at a particular time. Second, a *context network* is identified by recursively examining the activity of all neurons that led to each NR, a procedure referred to as a *backtrace* [11] (Fig. 3B). The first iteration of a backtrace identifies those neurons that were both anatomically connected to the NR neuron and active (above a threshold) at the previous time-step. The procedure can then be iterated as allowed by computational tractability. In general, a low iteration depth ensures the identification of the most salient neural interactions for a particular NR while avoiding a combinatorial explosion. Third, a Granger causality analysis is applied to assess the causal significance of each connection in the context network (Fig. 3C). In order to ensure robust statistical inferences, each connection is assessed over a time period considerably

[1]Recent methodological advances suggest that investigators soon may be able to characterize both anatomical and functional architectures in biological systems at microscopic scales. These advances include retrograde transneuronal transport of virus particles [19], metabolic markers for neuronal activity [20], and high-resolution optical imaging [21].

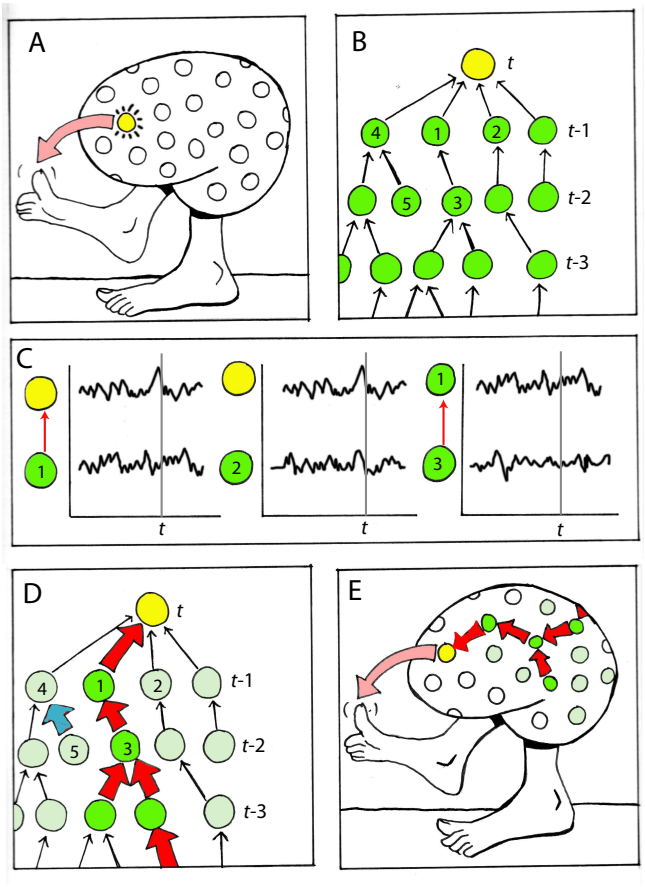

Fig. 3. Distinguishing causal interactions in neuronal populations. **A**. Select a *neural reference* (NR), i.e., the activity of a particular neuron (yellow) at a particular time (t). **B**. The *context network* of the NR corresponds to the network of all coactive and connected precursors, assessed over a short time period. **C**. Assess the Granger causality significance of each interaction in the context network, based on extended time-series of the activities of the corresponding neurons. Red arrows indicate causally significant interactions. **D, E**. The *causal core* of the NR (red arrows) is defined as that subset of the context network that is causally significant for the activity of the corresponding neuron [i.e., excluding both non-causal interactions (black arrows) and 'dead-end' causal interactions such as 5→4, indicated in blue].

longer than that used to identify the context network. Also, each connection is assessed separately; i.e., using a repeated bivariate design with correction for multiple comparisons. The resulting networks of significant Granger causality interactions are referred to as *Granger networks*. Last, the *causal core* of each NR is identified by extracting the subset of the corresponding Granger network consisting of all causally significant connections leading, via other causally significant connections, to the NR (Fig. 3D-E).

As reported in detail in [6], application of the above analysis to Darwin X involved selecting 93 NRs corresponding to bursts of activity in $CA1$ neuronal units at different time points spanning the learning process. Context networks for each NR were identified by iterating the backtrace algorithm for 6 time steps. Fig. 4 shows the context network, Granger network, and causal core for a representative NR. The causal core is strikingly small as compared to the context and Granger

networks, and is largely free from the intra-entorhinal interactions which dominate these other networks. These observations generalized to the remaining 92 NRs, suggesting that (i) even in large neural populations, only comparatively small subsets may be causally recruited at a given time for a given function, and (ii) trisynaptic and perforant pathways had greater causal influence on the selected NRs than did entorhinal interactions. Importantly, as reported in [6], causal cores could not in general be identified on the basis of synaptic strengths alone.

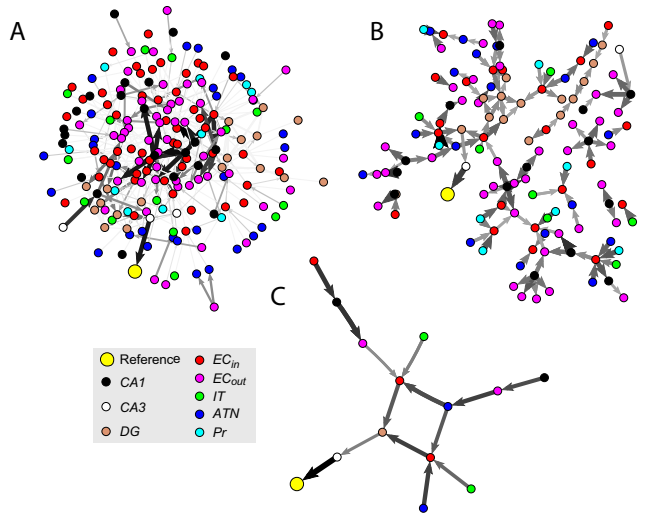

Fig. 4. **A**. The context network for a representative NR in Darwin X. The thickness of each line (and size of each arrowhead) is determined by the product of synaptic strength and presynaptic activity. **B**. The corresponding Granger network. Line thickness here reflects magnitude of the corresponding causal interaction. **C**. The corresponding causal core. Networks were visualized using the Pajek program (http://vlado.fmf.uni-lj.si/pub/networks/pajek/), which implements the Kamada-Kawai energy minimization algorithm. This figure is adapted from [6].

An interesting question that can be addressed with the above analysis is: How are causal interactions modulated during learning? Evidence indicates that learning can induce synaptic plasticity [22]. However, synaptic plasticity and behavioral learning operate over vastly different temporal and spatial scales [23] and, in part because of this, the precise functional contributions of synaptic plasticity to learned behavior have so far remained unclear.

Fig. 5 shows that causal cores reliably diminish in size as learning progresses; we have called this reduction in size *refinement* [6]. Because neither context networks nor Granger networks showed similar refinement during learning, causal core refinement in Darwin X may best be understood as arising from the selection of particular causal pathways from a diverse and dynamic repertoire of neural interactions. This is consistent with the notion [6] that synaptic plasticity may underlie behavioral learning via modulation of causal networks at the population level, and not by strengthening or weakening associative links between internal representations of objects and actions.

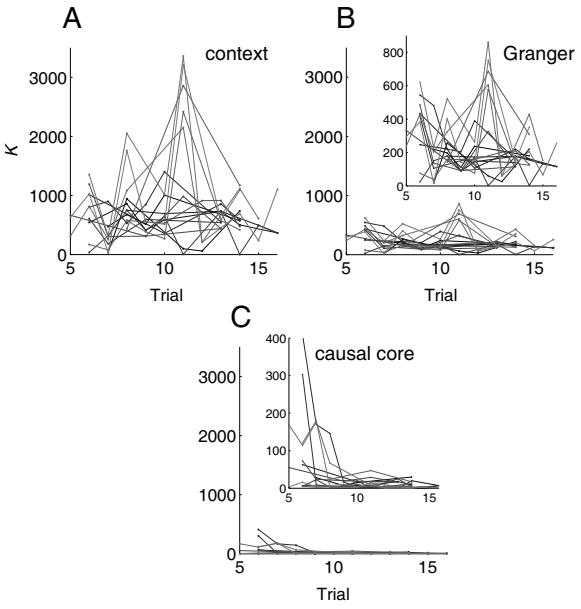

Fig. 5. **A**. Size of context networks as a function of trial number during learning, in terms of number of edges (K), for 15 different $CA1$ neuronal units. **B**. Sizes of the corresponding Granger networks. **C**. Sizes of the corresponding causal cores. Insets of panels **B** and **C** show the same data on a magnified scale. This figure is adapted from [6].

V. 'Causal density' and consciousness

The analyses described above demonstrate the usefulness of a causal network perspective for probing the functional connectivity of neural circuits and for analyzing the principles governing causal interactions in large neural populations. In this last section, we turn to the possible utility of such a perspective for characterizing the neural bases of consciousness.

A prominent theory of global brain function that has been applied to consciousness, the *theory of neuronal group selection* (TNGS), has proposed that consciousness is entailed by complex interactions among neural populations in the thalamocortical system, the so-called *dynamic core* [14], [15], [16], [24]. This proposal raises the question: How can these complex interactions best be characterized quantitatively?

A rewarding approach to this question is to consider phenomenology. A fundamental property of conscious scenes is that they are both *differentiated* (reflecting the discriminatory capability of consciousness; every conscious scene is one among a vast repertoire of different possible scenes) and *integrated* (reflecting the unity of conscious experience; every conscious scene is experienced "all of a piece") [14], [15], [16]. Therefore, a useful measure of complex neural interactions relevant to consciousness should reflect a balance between integration and differentiation in neural dynamics; this balance can be referred to as the *relevant complexity* of the system.

As we have argued previously [7], a useful quantitative measure of relevant complexity should also reflect the fact that consciousness is a dynamic process [25], and not a thing or a capacity; it should also take account of causal interactions

within a neural system and between a neural system and its surroundings, i.e., bodies and environments. To be practically applicable, a useful measure should also be calculable for systems composed of large numbers of interacting elements.

Several measures of relevant complexity have now been proposed, including 'neural complexity' (C_N) [26], 'information integration' (Φ) [27], and, most recently, causal density (c_d) [4]. All of these measures reflect in some way the balance between differentiation and integration of multivariate neural dynamics. In the present, we focus on causal density; a detailed comparative analysis of all three measures is provided in [7].

The causal density (c_d) of a network's dynamics measures the fraction of interactions among nodes that are causally significant [4], [7]. c_d is calculated as $\alpha/(n(n-1))$, where α is the total number of significant causal links observed, according to a multivariate Granger causality analysis, and n is the number of elements in the network. As reported in [7], it is also possible to calculate a 'weighted' version of c_d which takes into account the varying contributions of each causally significant interaction.

In terms of the criteria listed above, c_d naturally reflects a process because it is based on ongoing dynamics; it reflects causal interactions because it is based on an explicit statistical measure of causality. However, it is presently difficult to calculate for large systems because multivariate autoregressive models become difficult to estimate as the number of variables increases. It is possible that extended approaches based, for example, on Bayesian methods [28], may be able to address this practical limitation.

Does causal density capture aspects of relevant complexity? Fig. 6 shows a comparison of structural connectivity and causal connectivity for three example networks, along with the corresponding values of c_d. The dynamics for each network were generated using a mean-firing-rate neuronal model; each node received an independent Gaussian noise input. While both a fully connected network (having near-identical dynamics at each node) and a fully disconnected network (having independent dynamics at each node) have low c_d, a randomly connected network has a much higher value. These results support the notion that high values of c_d indicate that elements in a system are both globally coordinated in their activity (in order to be useful for predicting each other's activity) and at the same time dynamically distinct (reflecting the fact that different elements contribute in different ways to these predictions). High causal density is therefore consistent with high relevant complexity in that it reflects a dynamical balance between differentiation and integration.

It is important to recognize the limited role that a quantitative measure of neural dynamics can play within a scientific theory of consciousness. At the neural level, the relevant complexity of neural dynamics is likely to be multidimensional, involving spatial, temporal, and recursive aspects [7]. Because phenomenal states appear to exhibit a balance between differentiation and integration along each of these dimensions, a satisfying measure of relevant complexity must be sensitive to these aspects of complexity in neural dynamics; indeed,

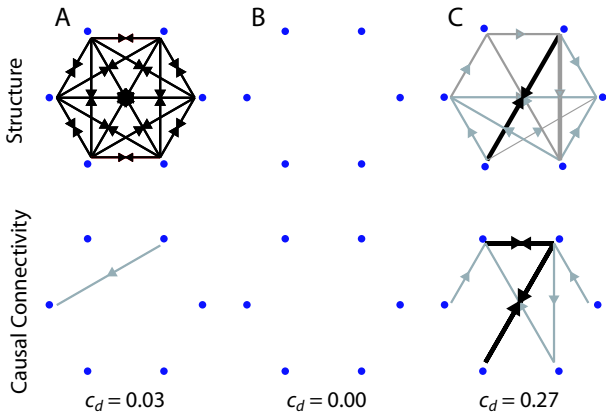

Fig. 6. Example simple networks (top row) and corresponding causal connectivity patterns (bottom row). **A**. Fully connected network. **B**. Fully disconnected network. **C**. Randomly connected network. Grey arrows show unidirectional connections and black arrows show bidirectional connections. The width of each arrow (and size of arrowhead) reflect the magnitude of the causal interaction. Corresponding values of causal density (c_d) are also given.

the simultaneous application of multiple measures may be required.[2] Like other measures of relevant complexity (i.e., C_N [26] and Φ [27]), c_d is best suited to measuring complexity in the spatial domain, although its basis in multivariate regression implies some integration over time.

More fundamentally, some aspects of consciousness are likely to resist quantification altogether. Conscious scenes have many diverse features, several of which do not appear to be readily quantifiable [14], [13], [7]. These features include subjectivity, the attribution of conscious experience to a self, and intentionality, which reflects the observation that consciousness is largely about events and objects.

In light of the above, it is clear that the quantitative characterization of relevant complexity can only constitute one aspect of a scientific theory of consciousness. It is worth noting that, unlike c_d or C_N, Φ has been proposed as an adequate measure of the amount consciousness generated by a system [27]; however, this claim is challenged not only by the above considerations but also by a demonstration that Φ can grow without bound even for a simple Hopfield-type network [7].

VI. Discussion

A causal network perspective provides a very general means of analyzing neural systems. As described above, causal networks can be used to trace functional pathways connecting sensory input to motor output, to distinguish causal interactions in large populations that lead to specific outputs, to explore how these interactions are modulated by synaptic plasticity, and to connect global features of neural dynamics to corresponding features of conscious scenes.

[2]Recursive complexity refers to the balance between differentiation and integration across different levels of description. The phenomenal structure of consciousness appears to be recursive inasmuch as individual features of conscious scenes are themselves Gestalts which share organizational properties with the conscious scene as a whole.

The Granger causality method for identifying causal networks can be contrasted with alternative techniques which require perturbation or lesioning of the studied system (e.g., [29], [30], [31]). The interpretation of causal inference based on such interventions is complicated by the fact that the studied system is either no longer intact (for lesions) or may display different behavior (for perturbations).

The identification of causal networks in neural systems involves no assumptions about whether these systems 'process information' or operate according to 'neural codes'. For example, causal cores are not information-bearing representations nor do they require explicit encoding and decoding. Instead, they are dynamic causally effective processes that give rise to specific outputs. This perspective can lead to different interpretations of commonly observed phenomena. For example, variability in neural activity [32] is often treated as an inconvenience and is minimized using averaging. According to the present view, however, such variability may indicate the existence of diverse and dynamic repertoires of neuronal interactions underlying causal core refinement during learning.

Although the experiments described in the present paper utilized a very simple implementation of Granger causality, recent methodological developments may lead to enhanced techniques better suited to the often non-linear and non-stationary time-series generated by embedded, embodied neural systems [33], [34]. When combined with network-theoretic and graph-theoretic techniques, these methods hold great promise for elucidating how neural systems function.

ACKNOWLEDGMENT

This research was supported by grant N00014-03-1-0980 from the Office of Naval Research, by DARPA, and by the Neuroscience Research Foundation. Many thanks to Dr. Gerald Edelman for useful comments and to Dr. Jeffrey Krichmar for his contribution to the design and construction of Darwin X. Thanks also to Bruno van Swinderen for help with Figure 1.

REFERENCES

[1] P. Churchland and T. Sejnowski, *The computational brain*. Cambridge, MA: MIT Press, 1994.

[2] R. deCharms and A. Zador, "Neural representation and the cortical code," *Annu Rev Neurosci*, vol. 23, pp. 613–47, 2000.

[3] C. Granger, "Investigating causal relations by econometric models and cross-spectral methods," *Econometrica*, vol. 37, pp. 424–438, 1969.

[4] A. Seth, "Causal connectivity of evolved neural networks during behavior," *Network: Computation in Neural Systems*, vol. 16, pp. 35–54, 2005.

[5] J. Krichmar, A. Seth, D. Nitz, J. Fleischer, and G. Edelman, "Spatial navigation and causal analysis in a brain-based device modeling cortical-hippocampal interactions," *Neuroinformatics*, vol. 3, no. 3, pp. 197–222, 2005.

[6] A. Seth and G. Edelman, "Distinguishing causal interactions in neural populations," *Neural Computation*, in press.

[7] A. Seth, E. Izhikevich, G. Reeke, and G. Edelman, "Theories and measures of consciousness: An extended framework," *Proceeding of the National Academy of Sciences, USA*, vol. 103, no. 28, pp. 10 799–10 804, 2006.

[8] C. Bernasconi and P. Konig, "On the directionality of cortical interactions studied by structural analysis of electrophysiological recordings," *Biological Cybernetics*, vol. 81, pp. 199–210, 1999.

[9] A. Brovelli, M. Ding, A. Ledberg, Y. Chen, R. Nakamura, and S. Bressler, "Beta oscillations in a large-scale sensorimotor cortical network: Directional influences revealed by Granger causality," *Proceedings of the National Academy of Sciences, USA*, vol. 101, no. 26, pp. 9849–9854, 2004.

[10] J. Krichmar and G. Edelman, "Machine psychology: Autonomous behavior, perceptual categorization and conditioning in a brain-based device," *Cerebral Cortex*, vol. 12, no. 8, pp. 818–30, 2002.

[11] J. Krichmar, D. Nitz, J. Gally, and G. Edelman, "Characterizing functional hippocampal pathways in a brain-based device as it solves a spatial memory task," *Proc. Natl. Acad. Sci. USA*, vol. 102, no. 6, pp. 2111–6, 2005.

[12] R. Morris, "Developments of a water-maze procedure for studying spatial learning in the rat." *Journal of Neuroscience Methods*, vol. 11, pp. 47–60, 1984.

[13] A. Seth, B. Baars, and D. Edelman, "Criteria for consciousness in humans and other mammals," *Consciousness and Cognition*, vol. 14, no. 1, pp. 119–139, 2005.

[14] G. Edelman, "Naturalizing consciousness: A theoretical framework," *Proceedings of the National Academy of Sciences, USA*, vol. 100, no. 9, pp. 5520–5524, 2003.

[15] G. Edelman and G. Tononi, *A universe of consciousness: How matter becomes imagination*. New York: Basic Books, 2000.

[16] G. Tononi and G. Edelman, "Consciousness and complexity," *Science*, vol. 282, pp. 1846–1851, 1998.

[17] J. Geweke, "Measurement of linear dependence and feedback between multiple time series," *Journal of the American Statistical Association*, vol. 77, pp. 304–13, 1982.

[18] E. Bienenstock, L. Cooper, and P. Munro, "Theory for the development of neuron selectivity: orientation specificit and binocular interaction in the visual cortex," *Journal of Neuroscience*, vol. 2, no. 1, pp. 32–48, 1982.

[19] R. Kelly and P. Strick, "Macro-architecture of basal ganglia loops with the cerebral cortex: use of rabies virus to reveal multisynaptic circuits," *Prog. Brain. Res.*, vol. 143, pp. 449–59, 2004.

[20] A. Konkle and C. Bielajew, "Tracing the neuroanatomical profiles of reward pathways with markers of neuronal activation," *Rev Neurosci*, vol. 15, no. 6, pp. 383–414, 2004.

[21] M. Raffi and R. Siegel, "Functional architecture of spatial attention in the parietal cortex of the behaving monkey," *Journal of Neuroscience*, vol. 25, pp. 5171–5186, 2005.

[22] R. Malenka and M. Bear, "LTP and LTD: An embarrasment of riches," *Neuron*, vol. 44, pp. 5–21, 2004.

[23] P. Drew and L. Abbott, "Extending the effects of spike-timing-dependent plasticity to behavioral timescales," *Proceedings of the National Academy of Sciences USA*, vol. 103, no. 23, pp. 8876–8881, 2006.

[24] A. Seth and B. Baars, "Neural Darwinism and consciousness," *Consciousness and Cognition*, vol. 14, pp. 140–168, 2005.

[25] W. James, "Does consciousness exist?" *Journal of Philosophy, Pyschology, and Scientific Methods*, vol. 1, pp. 477–491, 1904.

[26] G. Tononi, O. Sporns, and G. Edelman, "A measure for brain complexity: Relating functional segregation and integration in the nervous system," *Proceedings of the National Academy of Science (USA)*, vol. 91, pp. 5033–5037, 1994.

[27] G. Tononi, "An information integration theory of consciousness," *BMC Neuroscience*, vol. 5, no. 1, p. 42, 2004.

[28] A. Zellner, *An introduction to Bayesian inference in econometrics*. New York: Wiley, 1971.

[29] J. Pearl, *Causality: Models, reasoning, and inference*. Cambridge, UK: Cambridge University Press, 1999.

[30] G. Tononi and O. Sporns, "Measuring information integration," *BMC Neuroscience*, vol. 4, no. 1, p. 31, 2003.

[31] A. Keinan, B. Sandbank, C. Hilgetag, I. Meilijson, and E. Ruppin, "Fair attribution of functional contribution in artificial and biological networks," *Neural Computation*, vol. 16, pp. 1887–1915, 2004.

[32] A. Knoblauch and G. Palm, "What is signal and what is noise in the brain?" *Biosystems*, vol. 79, no. 1-3, pp. 83–90, 2005.

[33] M. Okatan, M. Wilson, and E. Brown, "Analyzing functional connectivity using a network likelihood model of ensemble neural spiking activity," *Neural Computation*, vol. 17, no. 9, pp. 1927–61, 2005.

[34] M. Kamiński and H. Liang, "Causal influence: Advances in neurosignal analysis," *Critical Reviews in Biomedical Engineering*, vol. 33, no. 4, pp. 347–430, 2005.