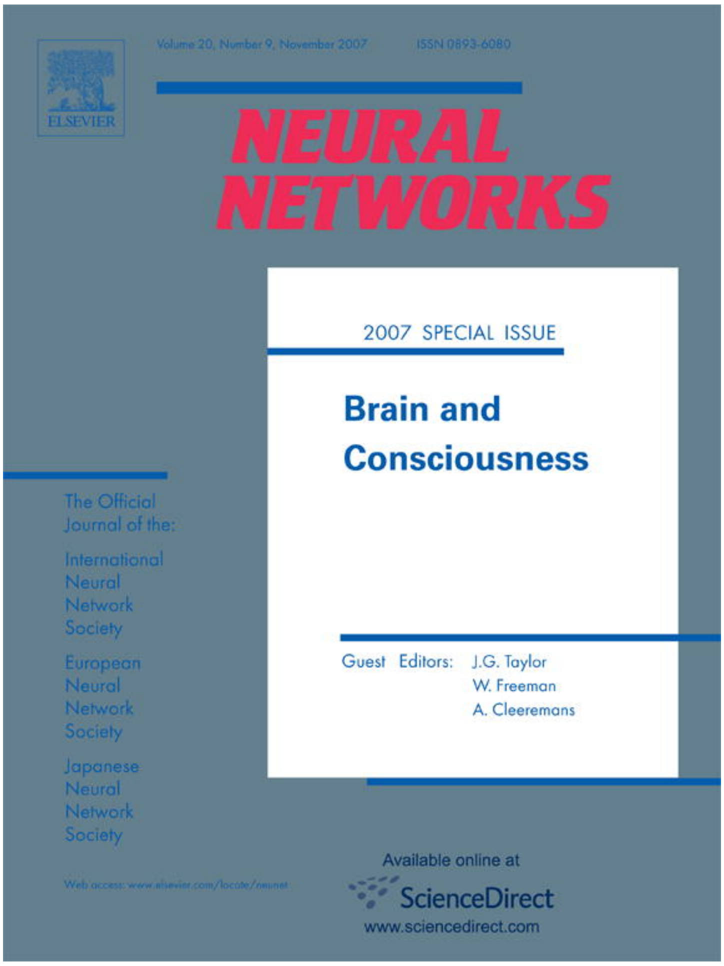

Volume 20, Number 9, November 2007 ISSN 0893-6080

ELSEVIER

NEURAL NETWORKS

2007 SPECIAL ISSUE

Brain and Consciousness

Guest Editors: J.G. Taylor
W. Freeman
A. Cleeremans

The Official Journal of the:
International Neural Network Society
European Neural Network Society
Japanese Neural Network Society

Web access: www.elsevier.com/locate/neunet

Available online at
ScienceDirect
www.sciencedirect.com

Available online at www.sciencedirect.com
ScienceDirect

Neural Networks 20 (2007) 1054–1056

Neural Networks

www.elsevier.com/locate/neunet

2007 Special Issue

Consciousness & the small network argument

Michael H. Herzog[a,*], Michael Esfeld[b], Wulfram Gerstner[c]

[a] *Laboratory of Psychophysics, Brain Mind Institute, École Polytechnique Fédérale de Lausanne (EPFL), Switzerland*
[b] *Department of Philosophy, University of Lausanne, Switzerland*
[c] *School of Computer and Communication Sciences and Brain-Mind Institute, École Polytechnique Fédérale de Lausanne (EPFL), Switzerland*

Abstract

The last decade has experienced a vivid enthusiasm to unravel the mystery of consciousness believed to be one of the major puzzles of human kind. We share this enthusiasm. Still, we feel that current models are incomplete suffering from a problem that we call the "small network argument".

Keywords: Neural networks; Consciousness; Recurrent connections; Modelling

Not even twenty years ago, consciousness has been widely considered a scientifically non-tractable issue. This view has greatly changed. In the last decade, a variety of brain processes has been proposed to account for consciousness. Typical examples are recurrent computations (e.g. Grossberg (1999) and Lamme (2006)), synchronized or oscillating neural activity (e.g. Bachmann (1994), Engel, Fries, Konig, Brecht, and Singer (1999) and Llinas, Ribary, Contreras, and Pedroarena (1998)), winner-take-all computations stabilized in resonance with the presynaptic neurons (Grossberg, 1999) or across the typical time span of working memory (Taylor, 1998), and closed loop action–perception processing (e.g. O'Regan and Noe (2001)). For example, recent physiological (Lamme, 1995) and psychophysical (e.g. Di Lollo, Enns, and Rensink (2000)) research has, indeed, suggested the importance of recurrent processing for consciousness, e.g. suppressing recurrent processing extinguished consciousness (e.g. Lamme (1995, 2006)). These findings are in good accordance with previous modelling studies proposing that all conscious states are resonant states[1] that stabilize over a time scale of a few hundred milliseconds (Grossberg, 1999; Taylor, 1998).

Whereas the theoretical and empirical results of these studies are of great importance, we propose that current models cannot fully account for consciousness because of a problem, we call the small network argument: For each of the above models, a very small neural network exists that fulfills the respective characteristics of the models but does not exhibit consciousness.

For example, two neurons, mutually interconnected, make up a recurrent system. Hence, these two neurons must create consciousness if recurrence is sufficient for consciousness (e.g. Lamme (2006)). Minimal models of winner-take-all computations require only three "competing" neurons which are fully connected to three presynaptic input neurons, plus potentially a single neuron controlling vigilance (Grossberg, 1999). Hence, such a network of seven neurons is sufficient to develop resonant states allowing learning (Grossberg, 1999) and working memory (Taylor, 1998). Analogously, if neural oscillations or synchrony are the main characteristics of consciousness, then, a group of three interconnected neurons firing in synchrony is conscious. Similarly, a thermostat, typically modelled as a single control loop between a temperature sensor ('perception') and an on–off switch for a heater ('action'), is a classical example of a perception–action device. It can be formulated as a two-neuron feedforward network with a sensory neuron connecting onto an output neuron controlling the heater switch.

If one does not want to attribute consciousness to such small networks other components are needed. For this reason,

[*] Corresponding author.
E-mail address: michael.herzog@epfl.ch (M.H. Herzog).

[1] Note that Grossberg emphasizes that the inverse statement 'is not yet asserted'.

0893-6080/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.neunet.2007.09.001

additional characteristics are, often implicitly, proposed to be necessary for consciousness. Typical examples are attention, cognition, the number of neurons, or the complexity of the network. Let us discuss these ideas in turn.

O'Regan and Noe (2001) combined their "sensomotor contingency" approach of perception with a work space like model (e.g. Newman, Baars, and Cho (1997)). Perception occurs in a closed loop of action and information processing. Consciousness emerges when attention and planning come into the play. For example, most of the normal car driving occurs in an automatic, unconscious mode. Conscious perception, e.g. of a traffic sign, emerges only if route planning and, hence, attention becomes important. Still, attention can be integrated within a small network just by adding one extra input arising from a second group of neurons (e.g. Hamker (2004))—containing potentially a very small number of cells.

Results from classical artificial intelligence have shown that invoking cognition does also not solve the small network problem either. For example, a basic calculator (or a universal computer running a program of only a few lines) outperforms most humans when it comes to computing the square root of 2 and it does so presumably without consciousness. Moreover, there are psychophysical indications that attention and cognition can occur without consciousness in human beings (e.g. Gaal, Ridderinkhof, Cornelisse, and Lamme (2007), Kiefer and Brendel (2006), Mattler (2005) and Scharlau and Ansorge (2003)).

Instead of attention and cognition, it is often proposed that consciousness emerges if a brain exceeds a certain number of neurons. However, let us suppose a model with a linear arrangement of neurons in which each neuron is connected to its neighbor to the right and left only. Given its simple connectivity, there is no obvious reason to assume that such a network, say, with 10^{10} neurons is more capable to create consciousness than its simplest version consisting of only three neurons. Hence, the sheer number of neurons alone is inadequate to overcome the small network argument. Therefore, other approaches state that a certain complexity of the connectivity of the network has to be met to yield consciousness. For example, Tononi and Edelman (1998) proposed to measure complexity by defining a functional cluster which is loosely connected to the rest of the network and has a rich repertoire of internal states. Still, even in this case we can construct a small network of, say, nine neurons, that meets the proposed complexity criterion.[2] Hence, the necessary ingredients of the theory of Tononi and Edelman (1998) can be implemented in a surprisingly small network.

We do not doubt that attention, recurrent computations, and complexity are important aspects to understand consciousness. However, we propose that these aspects are often trivially necessary rather than sufficient.[3] For example, often it is assumed that consciousness emerges not before several hundreds milliseconds after stimulus onset (e.g. Castiello, Paulignan, and Jeannerod (1991), Grossberg (1999), Libet, Gleason, Wright, and Pearl (1983) and Scharnowski et al. (2007), Taylor (1998)). Hence, given the short time constants of membranes of neurons, recurrent connections are obviously necessary to store and process the stimulus before consciousness is reached. Complexity is for sure of primary importance for consciousness because networks with the same number of neurons can create trivial as well as complex behavior depending on their connectivity.[4] Therefore, the important question is which kind of connectivity or which exact degree of complexity, determined with which mathematical norm, is sufficient for consciousness and why (but see Moody (2003))?

A rather different way to cope with the small network argument is to claim that, indeed, each small network has some kind of (almost vanishing) consciousness (e.g. Lamme (2006)). It is evident that such kind of "panpsychism" (cf. Globus (1976)) encounters serious problems as well. Consider, for example, two non-connected systems with three neurons each and each system having its own consciousness. What happens when these two systems are connected? Does one unified consciousness emerge, do the two consciousness' stay separate independent of each other (as proposed in split brain patients), or are there new coalitions of neurons making up new (micro)consciousness' (e.g. Zeki (2003))? In the later cases, there can be as many consciousness' at a given time as there are, for example, recurrent connections, synchronizations, or winner-take-all competitions. In short, it is by far not obvious why an increase in "conscious" elements yields one unity consciousness, we experience, and not many separate consciousness'.

We think that similar considerations hold also for other approaches not mentioned above linking consciousness, for example, to NMDA synapses (Flohr, 1992), to the biological tissue per se (Searle, 1992), and to metarepresentations (e.g. HOT: Rosenthal (1997); HOLT: Rolls (1997)).

In summary, we have argued that for each model of consciousness there exists a minimal model, i.e., a small neural network, that fulfills the respective criteria, but to which one would not like to assign consciousness. Appeals to additional aspects, such as the size or complexity of the network

[2] This network is organized in three clusters of three neurons each and weak connections between the three groups. The functional cluster index (Tononi & Edelman, 1998) takes high values if connections between the three neuronal groups are chosen arbitrarily small. Moreover, in a neuron model we may assign to each neuron n activity states (e.g., different firing rates or temporal patterns) and a suitable interaction dynamics on the time scale of hundreds of milliseconds. Suppose the following formal model of interaction: The state of each neuron is described by p bits, hence there are 2^p possible states. During the first time step of 100 milliseconds, the three neurons in a functional cluster agree on the first bit by a majority vote; in the following 50 milliseconds on the second bit, and during the final $200/2^p$ milliseconds on the pth bit. Hence, for any arbitrary p, the three neurons within the cluster will have agreed on p bits in less than 200 milliseconds leading to a high index of neural complexity.

[3] Precisely, they are part of a minimal sufficient condition for consciousness, whereby a condition is minimal sufficient iff none of its parts is redundant. However, they do not constitute a sufficient condition on their own. Something has to be added.

[4] Any kind of mathematical asymptotic behavior, i.e. stable fixed points, oscillations, and chaos, can be created with a network of three neurons where each neuron is described by a single nonlinear differential equation (e.g. Pasemann (2002)) and even by a single neuron model with three differential equations, as in the Hindmarsh and Rose (1984) model.

are obviously necessary but not sufficient to ‘explain’ how consciousness emerges. The above considerations should in no way hinder the exciting research on modelling consciousness. We do not claim that consciousness is a scientifically intractable problem in principle as some philosophers do (see notably Chalmers (1996) for an ontological and Levine (1983) for an epistemic argument). Quite to the contrary, we postulate that consciousness is accessible to natural science. We suggest to consider the small network argument as a benchmark any model of consciousness has to meet. In the past, benchmarks have helped strongly to structure research areas. For example, benchmark data sets have provided clear standards in machine learning. We suggest the small network argument to be a starting point for a set of benchmarks that may help to frame the problem of consciousness. However, at the current state, we do not see how computational approaches can escape the “small network argument” without appealing to unspecific or even mysterious forces.

Acknowledgment

We like to thank Markus Kiefer for useful comments.

References

Bachmann, T. (1994). *Psychophysiology of visual masking*. Commack, New York: Nova Science Publishers, Inc.

Chalmers, D. J. (1996). *The conscious mind. In search of a fundamental theory*. New York: Oxford University Press.

Castiello, U., Paulignan, Y., & Jeannerod, M. (1991). Temporal dissociation of motor response and subjective awareness. *Brain*, *114*, 2639–2655.

Di Lollo, V., Enns, J. T., & Rensink, R. A. (2000). Competition for consciousness among visual events: The psychophysics of reentrant visual processes. *Journal of Experimental Psychology*, *129*, 481–507.

Flohr, H. (1992). Die physiologischen Bedingungen des phaenomenalen Bewusstseins. *Forum fuer interdisziplinaere Forschung*, *1*, 49–55.

Engel, A. K., Fries, P., Konig, P., Brecht, M., & Singer, W. (1999). Temporal binding, binocular rivalry, and consciousness. *Consciousness and Cognition*, *8*, 128–151.

Globus, G. G. (1976). Mind, structure, and contradiction. In G.G. Globus, G. Maxwell, & I. Savodnik (Eds.), *Consciousness and the brain* (pp. 271–293). New York.

Grossberg, S. (1999). The Link between Brain Learning, Attention, and Consciousness. *Consciousness and Cognition*, *8*, 1–44.

Hamker, F. H. (2004). A dynamic model of how feature cues guide spatial attention. *Vision Research*, *5*, 501–521.

Hindmarsh, J. L., & Rose, R. M. (1984). A model of neuronal bursting using 3 coupled 1st order differential equations. *Proceedings of the Royal Society of London*, *B 221*, 87–102.

Kiefer, M., & Brendel, D. (2006). Attentional modulation of unconscious ‘automatic’ processes: Evidence from event-related potentials in a masked priming paradigm. *Journal of Cognitive Neuroscience*, *18*, 184–198.

Lamme, V. A. F. (1995). The neurophysiology of figure-ground segregation in primary visual cortex. *Journal of Neuroscience*, *15*, 1605–1615.

Lamme, V. A. (2006). Towards a true neural stance on consciousness. *Trends in Cognitive Science*, *10*, 494–501.

Levine, Joseph (1983). Materialism and qualia: The explanatory gap. *Pacific Philosophical Quarterly*, *64*, 354–361.

Libet, B., Gleason, C. A., Wright, E. W. J., & Pearl, D. K. (1983). Time of conscious intention to act in relation to onset of cerebral activities (readiness potential): The unconscious initiation of a freely voluntary act. *Brain*, *106*, 623–642.

Llinas, R., Ribary, U., Contreras, D., & Pedroarena, C. (1998). The neuronal basis for consciousness. *Philosophical Transactions of the Royal Society of London, B.*, *353*, 1841–1849.

Mattler, U. (2005). Inhibition and decay of motor and non-motor priming. *Perception & Psychophysics*, *67*, 285–300.

Moody, T. (2003). Consciousness and complexity. *Progress in Complexity, Information, and Design (PCID)*, *2.3.4*, 1–6.

Newman, J., Baars, B. J., & Cho, S. B. (1997). A neural global workspace model for conscious attention. *Neural Networks*, *10*, 1195–1206.

Pasemann, F. (2002). Complex dynamics and the structure of small neural networks. *Network*, *13*, 195–216.

O’Regan, J. K., & Noe, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences*, *24*(5), 939–973.

Rolls, E. T. (1997). Consciousness in neural networks? *Neural Networks*, *10*, 1227–1240.

Rosenthal, D. M. (1997). A theory of consciousness. In N. Block, O. Flanagan, & G. G”uzeldere (Eds.), *The nature of consciousness* (pp. 729–753). Cambridge, Massachusetts: MIT Press.

Scharlau, I., & Ansorge, U. (2003). Direct parameter specification of an attention shift: Evidence from perceptual latency priming. *Vision Research*, *43*, 1351–1363.

Scharnowski, F., Rüter, J., Jolij, J., Hermens, F., Kammer, T., & Herzog, M. H. (2007). Transcranial magnetic stimulation of early visual cortex reveals a window of integration of substantial duration. *Journal of Vision*, *7*(9), 1016.

Searle, J. R. (1992). *The rediscovery of the mind*. Cambridge, Mass: MIT Press.

Taylor, J. G. (1998). Constructing the relational mind. *Psyche*, *4*.

Tononi, G., & Edelman, G. M. (1998). Consciousness and complexity. *Science*, *282*, 1846–1851.

Van Gaal, S., Ridderinkhof, K. R., Cornelisse, S., & Lamme, V. A. F. (2007). Exploring the relationship between consciousness and cognitive control. Response inhibition can be triggered by masked stop-signals. *Journal of Vision*, *7*, 425a.

Zeki, S. (2003). The disunity of consciousness. *Trends in Cognitive Science*, *7*, 214–218.