

Alfredo Pereira Jr.
Jonathan C.W. Edwards
Dietrich Lehmann
Chris Nunn
Arnold Trehub
and Max Velmans

Understanding Consciousness

*A Collaborative Attempt to Elucidate
Contemporary Theories*

Nature Network Groups hosted an invited workshop on ‘Theories of Consciousness’ during the second semester of 2009. There were presentations by each of 15 authors¹ active in the field, followed by debate with other presenters and invitees. A week was allocated to each of the theories proposed; general discussion threads were also opened from time to time, as seemed appropriate. We (who had been participants in the workshop) offer here an account of the principal outcomes. It can be regarded as a contemporary, ‘state of the art’ snapshot of thinking in this field.

It should be said straight away that there was little general agreement in the workshop about what constituted the main problems, or how to address them. Explaining consciousness is particularly

Correspondence:

Alfredo Pereira Jr., São Paulo State University (UNESP), Institute of Biosciences,
18618-000, Botucatu-SP, Brasil e-mail: apj@ibb.unesp.br

-
- [1] Listed in alphabetical order: Bernard Baars, Ned Block, Jean-Pierre Changeux/Stanislas Dehaene, Axel Cleeremans, Jonathan Edwards, Vincent de Gardelle, Stuart Hameroff, Bjorn Merker, Chris Nunn, Alfredo Pereira Jr (discussing the work of Giulio Tononi and Christoph Koch), Susan Pockett, David Rosenthal, Anil Seth, Arnold Trehub and Max Velmans.

difficult, it soon became evident, because it involves so many different conceptual and scientific domains. For example, consciousness has both 'subjective' (first-person) and 'objective' (third-person) aspects. Relating the two aspects poses problems for scientific methodology.

Traditionally, philosophy and religion have been central to the issues, while contemporary scientific approaches depend on disciplines ranging from fundamental physics to sociology. Despite a range of differences over philosophical issues, most workshop participants appeared to agree that a reflexive form of Dual-Aspect Monism provides a reasonably satisfactory basis on which to proceed, taking conscious experiences and their physical (neural) counterparts to be complementary and mutually irreducible aspects of a single underlying dynamic.

Monism is usually contrasted with Substance Dualism, the view found for example in the writings of Plato and Descartes that, fundamentally, the universe is composed of two kinds of stuff, physical stuff and the stuff of soul, mind or consciousness. Reflexive Dual-Aspect Monism (Velmans, 1990; 2009) is a modern version of an ancient view that the basic stuff of which the universe is composed has the potential to manifest both physically and as conscious experience. In its evolution from some primal undifferentiated state, the universe, in a continuous or discontinuous fashion, differentiates into distinguishable physical entities, at least some of which have the potential for conscious experience. While remaining embedded within and dependent on the surrounding universe and composed of the same fundamental stuff, each human, equipped with perceptual and cognitive systems, has an individual perspective on both the rest of the universe and him/her self. In this sense, each human participates in a process whereby the universe differentiates into parts and becomes conscious in manifold ways of itself, making the entire process reflexive.

But how can we progress from philosophy to science? One method is to combine verbal or non-verbal reporting of conscious states with the execution of cognitive tasks and concomitant registers of brain activity (via fMRI or EEG for instance). Could it be taken for granted that report of consciousness is a reliable measure of its occurrence? A necessary connection of conscious (as opposed to automatic) processing with behaviour coordination has been proposed (Morsella, 2005). Making an analysis of this kind of approach, influential philosophers (e.g., Rosenthal, 2008) have questioned whether the connection is really necessary, since the bulk of behaviour control is actually

unconscious. From a practical point of view, too, accessing reports is fraught with problems, as in the case of ‘locked in’ patients, who cannot make behavioural reports; however, they can transmit their conscious thoughts by means of EEG signals, requiring analysis and linguistic interpretation by the observer (Birbaumer *et al.*, 1999). Therefore, our appraisal of the role of reportability for scientific research on consciousness is nuanced. On the one hand, first-person reportability as an index of consciousness is best understood as a potential or dispositional property that can be actualized or not, depending on particular circumstances. On the other hand, conventional EEG and more recent ‘mind-reading’ (Kay *et al.*, 2008) using fMRI provide a range of options for the identification of conscious processes, in addition to conventional first-person verbal reports, but the low spatial resolution of EEG and the low temporal resolution of fMRI limit the usefulness of these techniques.

The complexity of conscious content reflects the diversity of mental functions, such as affect, attention, perception, thinking, imagination, learning, memory, language, the sense of ‘Self’, decision-making and coordination of behaviour. Although each of these features appears to originate in the activities of specific brain mechanisms or ‘modules’, they all contribute content to what may be a common mechanism (the *caveat* here being that consciousness could be less unitary than is often assumed, from either ontogenetic or phylogenetic points of view). Many observations have shown that apparently similar functions can be either consciously experienced or remain unconscious (e.g., perceptual discrimination, as in the case of a mother woken from sleep by her infant’s faint noise and not by other people’s voices). One can distinguish high level information processing functions that are conscious, from the same functions in the brain that are non-conscious, e.g. the activation of semantic networks by consciously perceived linguistic stimuli contrasted with unconscious ‘semantic priming’ in masking experiments. This framework, made explicit in the ‘contrastive method’ advanced by Baars (1997), has motivated many experiments to dissociate conscious and unconscious aspects of mental functions, and corresponding brain activities. Those activities that correspond to conscious aspects have been called ‘neural correlates’.

While the search for brain correlates has been the dominant approach to the scientific study of consciousness, there are too many of these to usefully constrain our understanding of its underlying brain mechanisms. Arnold Trehub (an initiator of the Workshop) suggested the adoption of a ‘bridging principle’ to guide methodology,

namely that for any conscious event there is a corresponding analog in the biophysical state of the brain. One example of this principle is the correspondence between filling-in of conscious content (e.g. with a visual 'phantom') and the filling-in of physiological activity in the cortex (Meng *et al.*, 2005).

A concern that recurred throughout the workshop had to do with how best to define 'consciousness'. What operational definition is optimal when it comes to developing scientifically testable theories? If there is no agreed definition, would the choice depend on the theory being tested? If so, how can one choose the best theories? It was clear from the workshop that such foundational issues still need to be resolved.

There are, for example, theories concerning the biological substrates of consciousness. One question much discussed was: What kinds of brain states support the conscious state (as it is conceived in clinical practice)? Among proposed prerequisites for full consciousness are a functionally intact upper dorsal pons region (Plum and Posner, 1980) and a 'dominant' EEG frequency faster than 8 Hz (Niedermeyer and Lopes da Silva, 2005), although exceptions to such frequency patterns have been observed. It is commonly thought that neural synchrony plays an important role (Melloni *et al.*, 2007) and that slow cortical potentials may be involved (He and Raichle, 2009). There was a claim that consciousness *is* a pattern in brain electromagnetic fields (Pockett, 2000). Following a similar identity claim, Lehmann *et al.* (1998) have observed that spatial configurations of the EM fields change discontinuously over time, with quasi-stable spatial configurations called 'microstates' lasting about 100 milliseconds, and different microstates representing different thoughts. According to them, such microstates may be atoms of thought and emotion. Their sequence rule — the 'syntax' — is deviant in schizophrenia (Lehmann *et al.*, 2005). However, EEG patterns represent only one among a wide range of proposals for the neural basis of consciousness. Some theorists look for it in particular types of neural functional role, while others explore a range of ideas about the origins of consciousness in quantum biological processes.

A second issue concerns the nature of information patterns that compose conscious contents. The concept of information plays an important role in consciousness theories, due especially to the recently popular computer metaphor for brain function. In this context, several authors (such as Chalmers, 1996) attribute to informational concepts the epistemological role of bridging the gap between biophysical activity and phenomenal content. Different approaches to

information processing, from Shannon-Weaver mathematical to recent semantic and ecological concepts have been proposed. What kinds of information patterns are present in consciousness and how to describe them in a scientific model?

In addition to computational processes involving action potentials and synaptic weightings in neural nets, there have also been suggestions, as in the ‘microstate’ concept described above, that the informational content of consciousness is encoded in spatial configurations and amplitude, frequency and phase-modulated temporal patterns in local and/or global electromagnetic fields. More recently, the possibility that content is multiply encoded in local electrical fields in the dendrites of individual neurons, and made available to consciousness through coupling to local acoustic (phononic) modes, has been raised (Edwards, 2006).

The patterns can also be distinguished in terms of their origin: innate and/or learned (e.g. Cleermans, 2008, proposes that consciousness is learned). Among the innate aspects, a dynamic framework for the egocentric spatial representation of a volumetric surround (taken as one’s personal world) has been proposed (Trehub, 1991; Merker, 2009). Information dynamics can be understood as systems of attractors (Nunn, 2009), or on the basis of an excitatory threshold to be crossed (Del Cul *et al.*, 2009). Another prominent model proposes that consciousness reflects the level of information integration in the brain (Tononi and Koch, 2008).

A third issue concerns the existence (or not) of brain mechanisms that process available information patterns to generate conscious episodes. The now-classic theory addressing this issue is Global Workspace Theory (GWT), which claims that consciousness depends on a common workspace with output to and input from all cognitive processing modalities. There are other proposals that refer to specific brain mechanisms, such as the ‘retinoid’ system composed of autaptic cells, which constitute a neuronal embodiment of a global workspace (Trehub, 1991; 2007). At the quantum level of description, a well known hypothesis states that consciousness is based on coherent states generated in microtubules (Hameroff, 1998). Recently, a new proposal about brain mechanisms supporting conscious processes — the neuro-astroglial calcium wave — has appeared (Pereira and Furlan, 2009; Nunn, 2010).

An intensely debated issue was the identity of the conscious subject. Who is the entity that experiences conscious, cognitive and emotional contents? Is it correct to relate this entity with the psychological ‘Self’? Attempts to answer these intriguing questions include

single-cell consciousness, the idea that each neuron is a conscious entity (Edwards, 2005). A possibility implicitly assumed by many neuroscientists is that the Self is a brain network function. For Trehub (2007; 2009) the ‘Core Self’, corresponding to the origin of retinoid space, is an unchanging neuronal entity that anchors an ever-changing ‘Self Model’ (Metzinger, 2003). In cognitive sciences, the default position is that the conscious subject is the living individual (Pereira/Ricke, 2009). Finally, a recent proposal is that the conscious entity is like a traveling ‘pilot’ embodied in the activity of dendritic networks (Hameroff, 2010).

It would be fair to conclude that a science of consciousness may need a complex of theories to address different contributing factors requiring independent explanation. Further progress would then depend on an effective theoretical synthesis (as opposed to a conflict or competition) of theories that contribute to explain different aspects of studied phenomena. Any candidate for a unified theory of consciousness would need to be empirically validated by its ability to explain and predict conscious events, covering the above spectrum of relevant aspects raised by current theorists.

References

- Baars, B. (1988), *A Cognitive Theory of Consciousness* (New York: Cambridge University Press).
- Baars, B. (1997), *In the Theater of Consciousness: The Workspace of the Mind* (New York: Oxford University Press).
- Birbaumer, N., Ghanayim, N., Hinterberger, T., Iversen, I., Kotchoubey, B., Kübler, A., Perelmouter, J., Taub, E. & Flor, H. (1999), ‘A spelling device for the paralysed’, *Nature*, **398** (6725), pp. 297–298.
- Chalmers, D. (1996), *The Conscious Mind* (New York: Oxford University Press).
- Cleeremans, A. (2008), ‘Consciousness: The Radical Plasticity Thesis’, in: *Models of Brain and Mind: Physical, Computational, and Psychological Approaches. Progress in Brain Research 168*, eds. R. Banerjee & B.K. Chakrabarti (Amsterdam: Elsevier).
- Del Cul, A., Dehaene, S., Reyes, P., Bravo, E. & Slachevsky, A. (2009), ‘Causal role of prefrontal cortex in the threshold for access to consciousness’, *Brain*, **132**, pp. 2531–2540.
- Edwards, J.C.W. (2005), ‘Is consciousness only a property of individual cells?’, *Journal of Consciousness Studies*, **12** (4–5), pp. 60–76.
- Hameroff, S. (1998), ‘“Funda-Mentality”: Is the conscious mind subtly linked to a basic level of the universe?’, *Trends in Cognitive Science*, **2** (4).
- Hameroff, S. (2010), ‘The “conscious pilot” — dendritic synchrony moves through the brain to mediate consciousness’, *Journal of Biological Physics*, **36** (1), pp. 71–93.
- He, B.J. & Raichle, M.E. (2009), ‘The fMRI signal, slow cortical potential and consciousness’, *Trends in Cognitive Science*, **13**, pp. 302–309.

- Kay, K.N., Naselaris, T., Prenger, R.J. & Gallant, J.L. (2008), 'Identifying natural images from human brain activity', *Nature*, **452** (7185), pp. 352–355.
- Lehmann, D., Faber, P.L., Galderisi, S., Herrmann, W.M., Kinoshita, T., Koukkou, M., Mucci, A., Pascual-Marqui, R.D., Saito, N., Wackermann, J., Winterer, G. & Koenig, T. (2005), 'EEG microstate duration and syntax in acute, medication-naïve, first-episode schizophrenia: a multi-center study', *Psychiatry Research Neuroimaging*, **138** (2), pp. 141–156.
- Lehmann, D., Strik, W.K., Henggeler, B., Koenig, T. & Koukkou, M. (1998), 'Brain electric microstates and momentary conscious mind states as building blocks of spontaneous thinking: I. Visual imagery and abstract thoughts', *International Journal of Psychophysiology*, **29** (1), pp. 1–11.
- Melloni, L., Molina, C., Pena, M., Torres, D., Singer, W. & Rodriguez, E. (2007), 'Synchronization of neural activity across cortical areas correlates with conscious perception', *Journal of Neuroscience*, **27** (11), pp. 2858–2865.
- Meng, M., Remus, D.A. & Tong, F. (2005), 'Filling-in of visual phantoms in the human brain', *Nature Neuroscience*, **8** (9), pp. 1248–1254.
- Merker, B. (2009), 'Consciousness as Reality Format', unpublished paper.
- Metzinger, T. (2003), *Being No One: The self-model theory of subjectivity* (Cambridge, MA: MIT Press).
- Morsella, E. (2005), 'The function of phenomenal states: Supramodular interaction theory', *Psychological Review*, **112**, pp. 1000–1021.
- Nunn, C. (2007), *From Neurons to Notions: Brains, Minds and Meaning* (Edinburgh: Floris Books).
- Nunn, C. (2010), *Who was Mrs Willett? Landscapes and Dynamics of the Mind* (Exeter: Imprint Academic). To be published.
- Niedermeyer, E. & Lopes da Silva, F. (2005), *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields. 5th ed.* (New York: Wiley).
- Pereira Jr., A. & Furlan, F.A. (2009), 'On the role of synchrony for neuron-astrocyte interactions and perceptual conscious processing', *Journal of Biological Physics*, **35** (4), pp. 465–481.
- Pereira Jr., A. & Ricke, H. (2009), 'What is consciousness? Towards a preliminary definition', *Journal of Consciousness Studies*, **16** (5), pp. 28–45.
- Plum, F. & Posner, J.B. (1980), *The Diagnosis of Stupor and Coma, 3rd ed.* (Philadelphia, PA: FA Davis).
- Pockett, S. (2000), *The Nature of Consciousness: A Hypothesis* (San Jose, CA: Writers Club Press).
- Rosenthal, D. (2008), 'Consciousness and its function', *Neuropsychologia*, **46**, 829–840.
- Tononi, G. & Koch, C. (2008), 'The neural correlates of consciousness: An update', *Annals New York Academy of Science*, **1124**, pp. 239–261.
- Treuhub, A. (1991), *The Cognitive Brain* (Cambridge, MA: MIT Press).
- Treuhub, A. (2007), 'Space, self, and the theater of consciousness', *Consciousness and Cognition*, **16**, pp. 310–330.
- Treuhub, A. (2009), 'Two arguments for a pre-reflective core self: Commentary on Praetorius (2009)', *Consciousness and Cognition*, **18**, pp. 339–340.
- Velmans, M. (1990), 'Consciousness, brain, and the physical world', *Philosophical Psychology*, **3**, pp. 77–99.
- Velmans, M. (2009), *Understanding Consciousness. 2nd edition* (London: Routledge/Psychology Press).