ELSEVIER

Journal of Cognitive Systems Research 2 (2001) 241–244

Cognitive Systems RESEARCH

www.elsevier.com/locate/cogsys

Book review

Review of A Universe of Consciousness. How Matter becomes Imagination
G.M. Edelman & G. Tononi; Basic Books, New York, 2000; 266 pp; ISBN 0465013767

Action Editor: Stefan Wermter

Jelle van Dijk*, Gert-Jan Bleeker

Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

Received 1 November 2000; accepted 7 June 2001

New developments in neuroimaging are among the hottest topics in today's science. The mystery of consciousness is as old as human thought itself. Consequently, reference to both will interest many readers. Hence, a smart move was made by Gerald Edelman to expand his Theory of Neuronal Group Selection, or Neural Darwinism, by including both neuroimaging results and a hypothesis concerning the nature of conscious experience. The second smart move was to write the current volume together with colleague Giulio Tononi. The result, *A universe of consciousness, how matter becomes imagination*, is a well-written piece with a clear style and a balanced structure. The book takes a fundamental neuro-anatomical perspective with an emphasis on the complex dynamical organisation of the brain. Edelman & Tononi (E&T) claim that consciousness can actually be studied, and successfully explained scientifically. In doing so, the authors distantiate themselves from philosophically grounded treatments of the phenomenon. In fact, the book rejects the thesis that consciousness is, unlike other phenomena associated with brains, a *hard problem*: The thesis that no theory can ever be really about *consciousness* itself because the best theories can do is to discuss functional aspects of physical brains in conscious beings (cf. Chalmers, 1995).

The book is divided into six parts. In part I, the phenomenon of consciousness and the difficulty of the scientific problem are introduced, after which the writers progressively unfold their argument in the subsequent chapters, kicking off in part II by discussing some general characteristics of brain anatomy and organisation. Then, via the introduction of Neural Darwinism (part III) they go on to discuss some specific information-theoretical measures in part IV, upon which they finally arrive at the nature of consciousness as a neurodarwinistic *dynamic core*, covered in parts V and VI. In this review we will give a short summary of each of the parts, after which we will hopefully have gained enough material to end with a short discussion on the main question: Does this book actually present us with a viable route from matter to imagination, or do significant problems still block a successful journey?

Although consciousness has always been on the scientific agenda, until recently it has remained an issue touched only by philosophers. In part I, the authors claim that a fundamental limitation on philosophical efforts to discern the origins of consciousness arises from the presumption that the sources of conscious thought can be studied by thinking alone. Philosophers have described the *hard problem* of consciousness as being unsolvable: No theory will explain the generation of sensations, of phenomenal or experiential states, out of 'mindless' neuronal activity.

In psychology, introspection led to the search for conscious elements (sensations). Behaviourists left

*Corresponding author. Tel.: +31-24-352-1530.
E-mail address: jelle.vandijk@mpi.nl (J. van Dijk).

1389-0417/01/$ – see front matter
PII: S1389-0417(01)00033-X

consciousness untouched, whereas present-day cognitive psychologists use the concept of consciousness as a functional concept within an information-processing approach. Starting from the latter perspective neuroscientists try to find specific neural correlates of consciousness. None of these attempts has been very successful. The main problem is that we, conscious human beings, have to find a way to connect descriptions of subjective experiences with the actual subjective experiences themselves. Explaining and describing the objective physical processes involved in 'looking at a painting' comes nowhere near the subjective conscious experience of the looking itself. As McGinn (1989) put it: How do we turn the water of the physical brain into the wine of consciousness?

E&T try to work their way around this problem. Their strategy is focussed on the search for fundamental, general properties of consciousness that are shared by every conscious state. After identifying these, E&T examine what kind of neural processes actually explain these fundamental properties. In this way, consciousness becomes a collection of processes instead of an object. This allows consciousness to be studied scientifically.

In the final chapter of part I the general properties of consciousness are discussed. Integration, or unity, is one major property. Each conscious state is perceived as a whole. Another property is differentiation or informativeness: although each conscious experience is on the one hand united and undivided, it embeds great complexity considering the innumerable amount of possible conscious states we have. Additionally, conscious states are private in the sense that they are personal and inaccessible to others. In order to ground these properties, E&T present a number of examples from daily life and neuropsychological disorders. Each of these examples illustrates how privateness, unity and informativeness manifest themselves in conscious experience. The next challenge is to find actual brain events that cause these properties to arise.

If there is one thing to be learned in part II it is that the brain is unique in its organisation. Chapter 4 paves the road for the thesis that this uniqueness is critical to understanding consciousness. 100 billion nerve cells, a million billion synapses in the neocortex, extremely dense connectivity unique to each individual, in the words of E&T: 'we house a jungle in our heads'.

Three important systems are discerned. The first is the thalamocortical system, with local functional isolation in the form of specialised maps and, at the same time, massive integrative connectivity between multiple areas. This reciprocal interaction is sustained, according to E&T, by a process called *reentry*. Like in Edelmans earlier work, the concept of reentry will be the main tune played in the book (see also part III). Secondly, the thalamocortical system interacts with several appendages: cerebellum, basal ganglia, hippocampus, in long parallel uni-directional loops, circuits associated with functionally encapsulated routines such as motor planning and speech. The third system is called the 'value system': several nuclei in the brainstem and hypothalamus that project diffusely onto almost the entire cortex at once. It is believed the fanned projections of this evolutionary older system activate large areas of the cortex whenever something salient occurs, upon which plastic reorganisation of the cortex can take place.

According to E&T, the unique properties of the brain: connectivity, plasticity, categorisation, embodied values and reentrant dynamics are fundamentally unlike the digital computer, and the idea of the brain performing computations is hopelessly flawed.

The two main properties of consciousness, in E&T's conception, integration and differentiation, are discussed in the remainder of part II, with numerous references to psychiatric and neurological disorders and neuroimaging results. We are not sure whether the experimental results presented here really contribute to the power of the argument. For example, the claim that neural activation decreases as a result of skill learning is controversial since the peculiarities of neural plasticity are still heavily debated among cognitive neuroscientists at the moment (e.g. Recanzone (2000) presents evidence for both increases and decreases of neural activity during skill learning). Nevertheless, the main lesson to learn here is that wide-spread areas of the brain contribute to consciousness, and the system as a whole sustains both differentiation (functional specialisation) and integration (binding) of its activity. Although the brainstem is crucial for maintaining the *state* of consciousness, and activation of any specialised part of the cortex contributes to the *content* of a conscious experience, the obvious conclusion here is that consciousness will never be localised in the activity of a particular brain area.

Part III stresses the importance of Neural Darwinism. For those who have read the earlier books (e.g. Edelman, 1992) these chapters are mainly interesting to try and pin down precisely the proposed connection between reentry and consciousness. They will primarily be spelling out chapter 9, in order to assess whether that connection is made successfully. The argument goes somewhat like this: Reentry, a process that operates on various spatial and temporal scales, sustains reciprocal interaction between numerous specialised brain areas. It is always active, even without input, and it enables perceptual categorisation at the lower levels of organisation, as well as *global mappings* between online categorisations and memory at a higher level. Global mappings lead to a *remembered present*: a scene that signifies what the current categorisation has to do with what the organism *was doing earlier* (i.e. its historical and embodied memory). Here is where primary consciousness kicks in: 'The ability of an animal to connect events and signals in the world ... with its value-category memory system, to construct a scene that is related to its own history is the basis for the emergence of primary consciousness. ... The ability to construct a conscious scene is the ability to construct, within fractions of a second, a *remembered present* [p. 109, their emphasis]'.

The reader has to hold his breath for some chapters because the rest of the story follows later on. But what we learn for now is that the loops over the cortical appendages and fanned projections from the value system are both crucial players. The first consolidates memory; the second ensures that both memory and online categorisation are fundamentally embodied in nature. It is important to note here that E&T strongly reject cognitivism, where brain processes are seen as constituting computational processes on representations. The view presented is actually quite in line with an associationist story, but the extra features, reentry and the value system, can potentially overcome the classical problems of associationist theory (see Haselager, 1999).

By then we are still in need of a scientific method. Somehow the focus on differentiation and integration got lost a little in the previous chapters, and they are taken up again in part IV, which deals with them entirely. Chapter 10 identifies reentry as the mechanism that leads to integration. As the authors have stressed earlier, integration as well as differentiation, or informativeness, are the key-properties that make a scientific investigation of consciousness possible.

First, we are presented with a measure of *functional clustering*, which is defined as the 'integration' of a set of neural components, divided by the mutual information between that set and the rest of the brain. If a subset of the brain is highly integrated (meaning it has a great loss of entropy due to interactions among its elements) and at the same time the mutual information with the rest of the system is low (meaning this subset displays a relative autonomy) it is said to be a functional cluster. Functional clusters indicate integrative processes in the brain.

Next, the degree to which a neural element can differentiate among the states of the rest of the system is discussed. This is given by the *complexity* of the system. Complexity is taken to be a measure of the degree to which the brain has reached an 'optimal synthesis of functional specialisation and functional integration'. As an example case, epileptic seizures, accompanied by high synchronisation, nevertheless have low complexity because the number of possible differentiations of such a system is low. In sum, the definitions of complexity and integration are measurable quantities in combination with techniques like EEG, MEG, PET and fMRI. The question is, do these measurements actually tell us anything about the nature of consciousness?

In part V that question is addressed, with the introduction of yet another concept: the *dynamic core*. A dynamic core is a system of neural areas that interact more heavily among themselves than they do with the rest of the system. It is a system with high integration and high complexity. The reentrant activity on different levels – perceptual categorisation, global mapping – lead to the generation of a dynamic core within hundreds of milliseconds. The core permits the brain to select a dynamic state that connects multiple sources of information out of billions of possible other states. In this way the informational power of the brain is unlike any man-made machine up to date. A discrimination made by the core is what corresponds to the content of a conscious experience. Different parts of the brain can be jointly part of the core but the core is not restricted to any particular subset of the brain. However, there are areas such as those regulating blood pressure that are normally not part of it. Consequently, the information processed in such areas does not contribute to the content of conscious

experience. One intriguing hypothesis is that certain psychiatric syndromes may be associated with the existence of multiple dynamic cores within one brain.

We are now in the position to take up a question of qualia, 'the redness of red'. According to E&T, the concept of a dynamic core possesses all the general features of a conscious experience, such as privateness, differentiation and integration. The crucial difference between the conscious perception of 'red' and, say, a photodiode detecting 'red' is that the photodiode selects among only a few possible states, whereas the conscious system possesses a dynamic core that discriminates among billions of possible states.

Finally, in part VI we are challenged to push the argument further into the domains of language, the self, and ultimately into philosophic thought itself. Language is seen as crucial for the development of higher order consciousness, which permits being 'conscious of being conscious'. Ultimately we are 'prisoners of description', since 'being is not describing [p. 222]'. The reader is invited to determine for himself exactly what philosophical position is taken here.

A universe of consciousness provides the reader with an impressive amount of theoretical argumentation connecting ideas from neural modelling, philosophy, theoretical cognitive science and cognitive neuroscience. However, we have two main problems with the book. First, like the earlier work of Edelman, E&T do not often refer to other workers in the field, although many parallels exist. For instance people like Freeman, Grossberg, Kelso and also Erlhagen (e.g. Igel, Erlhagen & Jancke, 2000) seem to be worthy partners in the discussion.

However, the biggest problem, in our view, is that the nature of consciousness is still unexplained because two major gaps in the argument remain. First, how can we validate whether the measures of integration and complexity really detect the existence of a dynamic core (meaning the theoretical concept involving reentrance, perceptual categorisation, global mappings, personal historic memory and the resulting 'remembered present')? Isn't it likely that a host of other possible mechanisms exist in nature (clouds, economy, . . . ?) that also display high levels of integration and complexity, but do not contain reentrance etc.?

Secondly, even if the methodology accurately measures real dynamic cores, have we then thereby explained consciousness? We believe not, because the *hard problem* still stands as it is: The dynamic core tells us how a system can discriminate between billions of other states within a reference space, but why would this evoke conscious experience? Does it *have* to? E&T write on p. 167: ''The quale of the pure sensation of red *corresponds* to the discrimination that has been made among billions of other states within the same reference space [my emphasis]''. But 'corresponds to' is nagging here, because we still do not know what consciousness *is*. What is missing is the explanation of why it is unavoidable, given the theory, that such a discrimination *constitutes* consciousness.

As an alternative title, we would have called the book *''A universe of brain dynamics. How matter becomes cognition''*. On this subject the authors have a lot of exciting theory to share with us, with real possibilities of experimental research to it. To us, it doesn't really matter that consciousness stays beyond the limits of our imagination.

Acknowledgements

We wish to thank the members of the EEC reading group Nijmegen for the valuable discussions during the writing of this review.

References

Chalmers, D. J. (1995). The puzzle of consious experience. *Scientific American 273*, 80–86.

Edelman, G.M. (1992). *Bright air brilliant fire. On the matter of the mind*. New York: Basic books.

Igel, C., Erlhagen, W., & Jancke, D. (2001). Optimization of dynamic neural fields. *Neurocomputing 36*, 225–233.

Haselager, W.F.G. (1999). Neurodynamics and the revival of associationism in cognitive science. In: Riegler, A., Von Stein, A., & Peschl, M. (Eds.), *Does representation need reality?* New York: Plenum Press.

McGinn, C. (1989). Can we solve the mind-body problem? *Mind 98*, 349–366.

Recanzone, G.H. (2000). Cerebral cortical plasticity: perception and skill aquisition. In: Cazzaniga, M.S. (Ed.), *The new cognitive neurosciences*. Cambridge, MA. MIT Press.