Layout: T1 Standard SC | Book ID: 305843_1_En | Book ISBN: 978-3-642-37427-2
Chapter No.: 9 | Date: 20-4-2013 | Page: 1/16

Mechanism and Phenomenon of Consciousness

On Models and Ontology in Dennett and Edelman

Paolo Pecere

Abstract The neurological explanation of consciousness has become in the last decades a widespread field of research among neurobiologists and philosophers of mind. The development of experimental models of consciousness involves a parallel search for a suitable ontological background. Although most researchers share anti-dualistic and naturalistic ideas, there are controversial claims about the ontological interpretation of phenomenological data. After sketching some historical premises of this issue, the paper focuses on two case studies: Dennett's "multi-draft" model of consciousness, and Edelman's theory of consciousness, included in his "theory of the selection of neuronal groups". Edelman's theory turns out to provide a better solution to the open issues of contemporary research, since it avoids speculative hypotheses and dismissive attitudes, while leaving room for experimental and conceptual developments in a classical, "Newtonian" methodological style.

I will present some remarks about the use of mechanistic models in contemporary neurosciences and its ontological implications, focusing on two case studies: Dennett's radical program of a materialistic "explanation" of consciousness and Edelman's interpretation of his own neuroscientific model of consciousness. The discussion is best introduced by means of some introductory remarks about the Cartesian legacy in neuroscience, since contemporary issues about the neuroscientific explanation of consciousness—has it has been often recognized—still owe much to a Cartesian philosophical background.

P. Pecere (✉)
Università di Cassino e del Lazio Meridionale, Cassino, Italy
e-mail: paolopatch@yahoo.it

L. Magnani (ed.), *Model-Based Reasoning in Science and Technology*,
Studies in Applied Philosophy, Epistemology and Rational Ethics 8,
DOI: 10.1007/978-3-642-37428-9_9, © Springer-Verlag Berlin Heidelberg 2013

Layout: T1 Standard SC | Book ID: 305843_1_En | Book ISBN: 978-3-642-37427-2
Chapter No.: 9 | Date: 20-4-2013 | Page: 2/16

1 Mechanistic Models and Dualistic Metaphysics: A Cartesian Controversy and its Legacy

« I suppose the body to be nothing other than a statue or machine made of earth » ([1], XI, 120): this famous statement made by Descartes in his treatise *L'homme* (first published in 1662) largely influenced the study of the brain by connecting mechanistic physics with anatomical analysis. After seeing Descartes' book fresh off the press, Steno wrote in his *Discours sur l'anatomie du cerveau* (1665): « since the brain is a machine, we have no reason to hope to discover its design through means any different from those used for discovering the design of other machines. The only thing to do is what we would do with other machines, taking apart its components piece by piece and considering what they can do, separately and together » ([22], pp. 32–33). The use of a mechanistic model of the brain, of course, was limited in Cartesian philosophy by the metaphysical distinction between the essence of soul and the essence of body, which excluded the very possibility of explaining the higher mental faculties, and consciousness itself, by means of mechanistic physics. This claim was highly appreciated by thinkers such as Malebranche and Leibniz: the latter wrote, in a famous page of the *Monadology*, that « perception, and anything that depends on it, cannot be explained in terms of mechanistic causation » [est inexplicable par des raisons mecaniques], arguing that visiting the interior of a machine « would show you the working parts pushing each other, but never anything which would explain a perception » ([19], § 17, p. 609). Since the XVIIth century this claim has been contested by many thinkers—such as Spinoza and La Mettrie—who underscored the heuristic power of the mechanistic models for the understanding of the mind and presented dualism as a metaphysical prejudice and an impediment to scientific inquiry.

An anti-dualistic—and therefore anti-Cartesian—perspective has gained renewed attention in neurosciences of the second half of the twentieth century, as the developments in biology and the new techniques of brain-imaging have led to different attempts to explain the "mechanism" of consciousness, without resorting to dualistic ontological hypotheses on the mind. Anti-Cartesian chapters, in particular, are one of the common features of the main books on the theory of consciousness in the last 20 years, authored by both philosophers and neuroscientists who developed mathematical and/or mechanistic models of consciousness and its different properties: think of Gerald Edelman and Antonio Damasio, Patricia Churchland and Daniel Dennett. A different, sympathetic judgment about Descartes's legacy has been formulated by French leading neuroscientist Jean-Pierre Changeux, who considers Descartes as a major forerunner of any successive physical explanation of the brain functions. Changeux adheres to the old fashioned—yet quite questionable—historiographical idea that Spinoza and La Mettrie represent the straightforward development of Descartes' mechanistic program, whose implications were materialistic from the outset ([6], pp. 47–54). Changeux himself claimed that the hypothesis of a physical explanation of the mind

Layout: T1 Standard SC | Book ID: 305843_1_En | Book ISBN: 978-3-642-37427-2
Chapter No.: 9 | Date: 20-4-2013 | Page: 3/16

(including consciousness) is the only heuristically positive option for neuroscientific research. As he put it in his programmatic book *L'homme neuronale*, there is no way for neuroscience but to assault the « Bastille of mind » ([5], p. 210).

In spite of any polemical accent, it should be recognized that both (a) the idea of mind and (b) the idea of mechanical explanation, that form the background for contemporary anti-Cartesian programs, owe much to the metaphysical foundation of modern science of nature provided by Cartesian philosophy.[1]

(a) The contemporary « problem of consciousness », as the quest for the neurobiological explanation of the most general qualitative feature of experience, usually presupposes Descartes' identification of the mind with « thought » , considered, in turn, as « everything which takes place in us so that we are conscious of it » ([1], VIII, p. 7). This problem, to be sure, would make no sense on the background of—say—Aristotelian hylemorphism, with its threefold soul as the "substantial form" of life and thinking, for here there is no gap to be further explained between matter and mind. Indeed, Descartes' list of biological phenomena than can be explained by means of a purely mechanistic account (see [1], XI, p. 202; [1], VII, pp. 229–230), includes all the functions of peripatetic vegetative and sensitive soul, which, in turn, correspond to the contemporary « neurological unconscious » , as the set of genetically or empirically stored abilities. Descartes' « mind » corresponds on the other hand to the conscious sensitive, imaginative and intellectual perceptions, which are precisely the phenomena investigated by contemporary theories of consciousness.

(b) The explanatory models developed by contemporary neuroscience are grounded on neurons and their physico-chemical activities, and as such they reflect the metaphysical distinction of matter (as lifeless extension) from mind operated by Descartes, while rejecting at the same time the very existence of a separate immaterial soul. This produces the need for an alternative explanation of consciousness and voluntary activity.[2]

On the whole, one can say that Cartesian ideas of matter and mind are essential for the very formulation of research programs in contemporary brain and cognitive sciences. This is still true today, as this conceptual heritage is more and more acknowledged and considered controversial by several leading researchers in both philosophy of mind and neurosciences: to dispose of mind-matter dualism is

[1] The presence of Cartesian ideas in philosophy of mind and neuroscience has been noticed several times in the twentieth century. This has been often considered as a starting point for philosophical criticism, which has been advanced by quite different perspectives (just think of Ryle and Heidegger). Most recently the dependence of twentieth century neuroscience on Cartesian dualism has been investigated by Maxwell Bennett and Peter Hacker. These authors consider the Cartesian attribution of mental properties to the soul as the exemplar model of the "mereological fallacy" of attributing mental faculties to the brain, which would be widely present in neuroscience ([2], pp. 43–44, 68, 160–161). On this part-whole problem see § 5 below.

[2] According to Descartes, we have « clear and distinct » ideas of both the separate existence of the immaterial soul and the action of the soul on the body. The latter's evidence therefore is not to be disputed or further analyzed. For a penetrating account see Garber [18], pp. 168–188.

Layout: T1 Standard SC | Book ID: 305843_1_En | Book ISBN: 978-3-642-37427-2
Chapter No.: 9 | Date: 20-4-2013 | Page: 4/16

indeed an ontological ideal that plays a crucial role in recent research on the neural correlates of consciousness.

In front of this complex Cartesian legacy, which here I cannot examine more in detail,[3] a series of questions arises: can the models of consciousness developed in contemporary neurosciences be considered as steps towards a reduction of consciousness to a mechanical process (that is, an ontological reduction of consciousness to matter)? Or do they play a heuristic role in the search for a ontologically different, non-materialistic theory? Which is, on the other hand, the methodological role of the very phenomenological evidence about conscious thinking that led Descartes to postulate metaphysical dualism? In contemporary philosophy and neurosciences these issues turn out to be controversial, as I will try to show by analyzing two different and contrasting cases.

2 Dennett: Mechanical Hypothesis and Explanation of Consciousness

Daniel Dennett, in his book *Consciousness explained* [9], sets out an explanatory hypothesis about consciousness which is connected to a « naturalistic-mechanical » ontology. He presents his view as opposed to the « reactionary » claim of those (such as Noam Chomsky, Thomas Nagel, Colin McGinn) who deny the possibility of a naturalistic-mechanical explanation of consciousness, and traces this view back to the Cartesian Age and to Leibniz's mistaken conflation of an epistemic problem with an ontological judgment.[4] Dennett's commitment to the naturalistic program is presented as a heuristic consequence of the fact that mind–body dualism is an impediment to scientific research and encourages the anti-scientific claim that consciousness is a « mystery ». According to Dennett, indeed, the introduction of mental properties in the description of conscious processes provides no theory at all, or, as he pungently puts it: « accepting dualism is giving up ».[5]

[3] It must be observed that Descartes himself devoted a substantial part of his work to discussing the unity of body and soul in its metaphysical, medical and ethical aspects. This fact is usually not recognized in contemporary criticism against Cartesian dualism. For an overview of this aspect of Descartes' philosophy, including an appraisal of its seminal role in grounding psycho-physical explanations in medicine, see Voss [29], pp. 186–196.

[4] For the latter argument see Dennett [11], pp. 1–10. I will elaborate on this problem in § 5 below.

[5] Dennett [9], p. 37: « There is a lurking suspicion that the most attractive feature of mind stuff is its promise of being so mysterious that it keeps science at bay forever. This fundamentally antiscientific stance of dualism is, to my mind, its most disqualifying feature, and it is the reason why in this book I adopt the apparently dogmatic rule that dualism is to be avoided at all costs. It is not that I think I can give a knock-down proof that dualism, in all its forms, if false or incoherent, but that, given the way dualism wallows in mystery, accepting dualism is giving up ».

Layout: T1 Standard SC | Book ID: 305843_1_En | Book ISBN: 978-3-642-37427-2
Chapter No.: 9 | Date: 20-4-2013 | Page: 5/16

Dennett therefore supports an explanatory theory of consciousness « within the framework of contemporary physical science » ([9], p. 40). By contemporary physical science Dennett considers standard physical science, and does not consider speculative conjectures such as Penrose's and Chalmers' about the possibility of new physical theories. Therefore, Dennett's program is to eliminate « mind-stuff » and to explain consciousness as a product of normally considered bio-physical processes. Such a program, of course, is not altogether new: Dennett recasts reductionist and physicalistic ideas of twentieth century philosophy of mind, drawing on the most recent tools of cognitive science and connectionism.[6]

According to Dennett « human consciousness is itself a huge complex of memes (or more exactly, meme-effects in brains) that can best be understood as the operation of a "von Neumannesque" virtual machine implemented in the parallel architecture of the brain that was not designed for any such activities » ([9], p. 210). Dennett's hypothesis is constructed by drawing on contemporary developments in computer science (Von Neumann), linguistics (Levelt) and evolutionary biology (Dawkins). It considers consciousness as a « virtual machine » implemented in the brain; its ability to represent and express meanings is not subject to a central control (a central « Meaner »), but rather depends on a subconscious competition of « multiple drafts » , produced by parallel processes in different regions of the brain, whose resolution, that eventually leads to speech acts, depends on pragmatic criteria. The communication of meanings, in turn, corresponds to the ability to share « memes » in cultural networks, one of them being the very idea of the Self.

One of the most striking aspects of Dennett's hypothesis is the ontological denial of qualia, which depends on the philosophical criticism of the illusory contents of phenomenology. Contrasting the very idea of a quale, as a supposedly irreducible element of experience, Dennett argues that this can indeed be analyzed and "explained away" in terms of information and belief, and therefore it only exists in a fictional sense, rather than in a natural sense:

> « Heterophenomenological objects—i.e. qualia—are, like centers of gravity or the Equator, abstracta, not concreta. They are not idle fantasies, but hardworking theorist's fictions » ([9], pp. 95–96)
>
> «The heterophenomenology exists—just as uncontroversially as novels and other fictions exist. People undoubtedly do believe they have mental images, pains, perceptual experiences, and all the rest, and these facts—the facts about what people believe, and report when they express their beliefs—are phenomena any scientific theory of the mind must account for » ([9], p. 98).

[6] There are already several introductions to Dennett's theory of consciousness. For an overview see Schneider [25] and the brief critical assessments by Andrew Brook and Paul Churchland in Brook/Ross [4], pp. 41–63, 64–80. For a useful historical survey of physicalistic and anti-physicalistic trends in twentieth century philosophy of mind see Moravia [21], which does not cover contemporary naturalism such as Dennett's. For a more up to date account on contemporary issues see Velmans/Schneider [28] and McLaughlin [20].

Dennett rejects the label of "eliminativism", insisting that his point here is to reconsider mind without mystery.[7] Given this important clarification, one still has to recognize that Dennett eliminates mind as a separate property or substance, reducing it to a fiction and an object of belief. But this does not mean that Dennett only wants to construe consciousness as being a matter of language: belief itself is a material brain process and therefore consciousness is indirectly inserted in a physical background. This crucial point can be highlighted by considering how Dennett's program is rooted in Ryle's and Wittgenstein's philosophy of mind, and at the same time connects the latter's criticism of mind to a brand new materialistic *pars construens*: Dennett's ontological commitment with the machine model strictly depends on the philosophical project to connect a Wittgensteinian anti-metaphysical criticism of language, and in particular of private feelings, with an evolutionary and materialistic theory of mind framed within the tradition of computational cognitive science.[8]

In front of this bold theoretical claim, it is interesting to observe that what Dennett presents is « just the beginning of an explanation » ([9], p. 455). To be sure, according to Dennett the task of philosophy is to show whether such a theory is possible or impossible ([9], p. 41). Therefore he defends the possibility of a non-dualistic hypothesis, without entering the details of its realization. « All I have done, really, is to replace one family of metaphors and images with another […] It's just a war of metaphors, you say—but metaphors are not "just" metaphors; metaphors are the tools of thought » ([9], p. 455). One may wonder, then, whether the materialistic reduction of consciousness in terms of material processes can be construed as a heuristic maxim, rather than as a fully grounded ontological commitment.

3 Edelman's TSNG and the Role of Phenomenal Consciousness

It is very instructive to compare Dennett's variously and strongly philosophically oriented conjecture with Edelman's theory of the selection of neuronal groups (TSNG), which presents a quite different interpretation of the role of phenomenology in the scientific description of consciousness. Edelman's theory—first fully articulated in *Remembered Present* [12] and later in a number of books such as *A*

[7] « Am I an eliminativist? I am a deflationist. The idea is to chip the phenomena of the mind down to size, undoing the work of inflationists who actively desider to impress upon themselves and everybody else just how supercalifragilisticexpialidocious consciousness is, so that they can maintain, with a straight face, their favourite doctrine: The Mind is a Mystery Beyond All Understanding » ([10], pp. 369–370).

[8] On this point it is very instructive to consider Dennett's critical exchange with Maxwell Bennett and Peter Hacker, who defend a different development of Wittgenstein's ideas ([3], for Dennett's view see in part. pp. 77–89).

Universe of Consciousness [16]—includes probably the most elaborated evolutionary and anti-dualistic theory of consciousness in contemporary neuroscience. It is grounded on three empirical principles[9]:

(1) **Developmental selection**, as the formation of the gross anatomy of the brain, which is partly controlled by genetic factors, but involves a high degree of individual variation in the neural connectivity;
(2) **Experiential selection**, a continuous process of synaptic selection, occuring within the diverse repertoires of neuronal groups. This process may strengthen or weaken the connections among groups of neurons and it is constrained by value signals that arise from the activity of the ascending systems of the brain, which are continually modified by successful output;
(3) **Reentry**. The ongoing recursive dynamic interchange of signals that occurs in parallel among connected brain areas, and which continuously coordinates in time and space the activity of their maps. Edelman considers a massive presence of reentry as a distinctive feature of human brain.

On this background Edelman develops his hypothesis about the neural correlates of consciousness. In Edelman's model, consciousness depends at any given moment on the activity of different and distributed groups of neurons, which form the so-called « dynamical nucleus ». The dynamical nucleus is defined as a « functional cluster » of neurons, connected by reentrant interactions. Consciousness, on the other hand, is defined as the « ability to construct a scene » and operate multidimensional « discriminations » inside this scene.

The task of the theory, now, is to explain the emergence of consciousness, as it is phenomenally given: this includes properties such as unity, qualitativity, temporal ordering, intentionality ([14], pp. 119–120) Edelman argues that the phenomenology of consciousness can be connected with the underlying neural processes by means of different features of the latter's integration and differentiation of information. The constantly changing and integrating components of the dynamical nucleus, for instance, « correspond » to the changing contents of conscious experience and their temporal ordering: the dynamical connection of « value-category memory » and « perceptual categorization » first produced consciousness as a « remembered present » ([14], p. 55). The selective integration of different cortical maps accounts for the constructive aspect of consciousness (closure, filling of gaps, Gestalt effects). A quantitative measure of functional integration and differentiation helps to connect this hypothesis with mathematical models and to design experimental tests. The multifarious afference of sensory information and its mnemonic modulation account for the rich qualitative contents of experience.

[9] See Edelman [14], pp. 39–41. I will consider here the most recent expositions of the theory (starting from Edelman/Tononi 2001), which probably take into account some philosophical criticism of previous expositions. See below note 10.

In Edelman's model, on the whole, properties of the neural network correspond to phenomenal properties: here lies their "explicative" value. Edelman asserts very clearly that his theory (and any explanatory theory in general) cannot ever reproduce qualitative experience and that postulating, in this sense, the reduction of conscious experience to neural activity is a « category mistake ».[10] This is indeed a crucial point of his theory, which must be carefully analyzed. One has to separate the impossibility to reduce and therefore « eliminate » conscious experiences by means of neurological description from the « methodological inability » of present neuroscience to provide such a description, which is theoretically possible.[11] Even if we had a perfect neurological description of the immensely complex neural interactions (that is, to put it in Searle's terms, even if we knew « exactly how » reentrant mechanisms cause conscious states: see footnote n. 10), then we would not have « eliminated » or « reduced » (or, to put it in Dennett's terms, « explained away ») the conscious experience, as the natural way that enables us to be informed about our interaction with the world.

4 Phenomenon and Mechanism of Consciousness: A Comparison Between Dennett and Edelman

The exposition of Edelman's theory has already introduced the strong difference between his own and Dennett's program. First, Edelman's elaboration on selection theory is considerably different from Dennett's. Whereas in Dennett selection occurs among possible speech acts, on the "software" level, in Edelman the selection of the fittest populations of neurons is a fundamental feature of the brain connectivity and results in the plasticity of the neural architecture itself: in human

[10] Edelman [14], p. 125. This is possibly a reply to critical remarks advanced by John Searle with regference to Edelman's previous books *Remembered Present* and *Bright Air, Brilliant Fire* [12, 13]. Searle considers Edelman's theory as « the most thorough and profound attempt that I have seen in the neurobiological literature to deal with the problem of consciousness ». Nonetheless, he considers Edelman's theory unsatisfactory, because it does not explain how qualia are produced by the neural activity: « Assuming that we understand how the reentrant mechanisms cause the brain to develop unconscious categories corresponding to its stimulus inputs, how exactly do the reentrant mechanisms also cause states of awareness? One might argue that any brain sufficiently rich to have all this apparatus in operation would necessarily have to be conscious. But for such a causal hypothesis the same question remains—how does it cause consciousness? And is it really the case that brains that have these mechanisms are conscious and those that do not are not? So the mystery remains » ([26], pp. 48, 50). Searle's essay is a revised version of a review in « The New York Book Review », November 16, 1995.

[11] See e.g. Edelman [15], p. 145: « Indeed, at present, because we lack the means of fully detailing the hyperastronomical interactions of core neurons, C [the conscious system] provides the only indicator we have of any overall core state, C' [the neural system]. Indeed, our methodological inability to reduce to cellular or molecular terms the mental or conscious events accompanying fields such as ethics and aesthetics that emerge when we speak "C language" to each other should not be construed as arising from the existence of some radically inaccessible domain ».

brains, the "hardware" level cannot be separated from the "software" level. Second, Edelman is more cautious about the possibility of neurological explanations of single phenomenological data. The attempt to develop a physico-mathematical model of the brain processes results in the admission that the only evidence that can be mastered of such a complex physical system is—at present—statistical: the dynamical nucleus is defined by means of a measure of « neural complexity », grounded on the statistical theory of information.[12] But there is no evidence, in Edelman's works, that a more advanced theory will be able to provide a more finely grained mechanical description of single qualia, for the latter are inserted in the unitary and multidimensional conscious scene corresponding to the dynamical nucleus.

On the whole, though Edelman intends to « complete Darwin's program » ([14], pp. 1–3) by naturalizing mind in terms of biological evolution, his global biological approach provides a quite different way to naturalism than Dennett's. Two global features of the nervous system—plasticity of the brain and complexity of neural interactions—support Edelman's conclusion that phenomenal consciousness is a unique means of understanding the human mind, that was developed in the evolution of the human organism in order to represent the individually different and highly complex brain processes. Edelman considers this conclusion to be in direct opposition to the computer science model of the brain, and this rebuttal implies the joint rejection of any eliminationist program in neuroscience, including those that are built on computational metaphors. Contrary to the computational model, the phenomenal content of mind is not the single-channel output reduction of a parallel process of elaboration of data (or even, as Dennett puts it, a misleading construct of folk beliefs); phenomenal consciousness peculiarly expresses « complex discriminations » produced by the parallel activity of the brain, in order to put them at work by interacting with the ambient: « qualia » —as Edelman repeatedly underscores— « are these discriminations » , and therefore they exist ([14], p. 70). Indeed, the very « logical » model of mind, as a set of rules designed to perfectly decipher codified sense-data, is opposed by Edelman to his own interpretation of the fundamental process of mind, the « recognition of configurations ».

To be sure, Dennett and Edelman share some crucial views about neuroscience. They both consider a « mechanistic » model as heuristically fundamental for the sake of a scientific theory of consciousness, and support an ultimately naturalistic ontology, at least insofar as consciousness is identified with a neural « process » and does not require the position of any immaterial being. Both Dennett and Edelman, moreover, are active in the field of A.I. and consider the development of thinking artifacts as a crucial enterprise in order best to understand and possibly to

[12] This theory is presented in Edelman/Tononi [16], pp. 125–138. Since this section of the book contains mainly Edelman's and Tononi's technical work on the measurement problem I do not analyze it in details. See Tononi [23].

reproduce human thinking.[13] Nonetheless there are philosophically crucial differences in the way mechanism and phenomenon are related in their different theories, which we can summarize by distinguishing the (a) ontological from the (b) methodological point of view:

(a) According to Dennett, the phenomenon of consciousness (as the representation of qualia) is just an illusory, methodologically misleading and ontologically empty content, that has to be explained away. According to Edelman, it is the only means to represent the « complex differentiations » operated by the brain processes and indeed it exists in human beings as a result of evolution and cannot be dismissed as an illusory theoretical construct.
(b) This ontological difference involves a substantive epistemological difference: whereas for Dennett mechanical models are initially introduced as a metaphor but eventually, being the only promising scientific description of the data, they have to correspond to a true description of what there is—mind is a property of a complex machine—in Edelman the phenomenal content of human consciousness adequately expresses a fundamental feature of brain processes themselves. This shift can be usefully expressed in terms of models and ontology: whereas in Dennett mechanistic (computational) models of consciousness reflect a materialistic ontology, which does not leave room for any genuinely phenomenological property, in Edelman the model of the brain network reflects the phenomenological properties of consciousness itself without excluding the latter's existence. Indeed, one could even say that for Edelman consciousness itself, due to its epistemic role for human beings, is a kind of "natural model" of highly complex brain processes.

Now, going back to Descartes' legacy, one may wonder whether this distinction of consciousness from brain processes amounts to a new dualism. Regarding the distinction of consciousness from the corresponding dynamical process Edelman writes:

> « the dynamic structural origin of properties, even conscious properties, need not resemble the properties it gives rise to: an explosion does not resemble an explosive » ([14], p. 63).

Edelman's terminology is not very strict about the relation between brain and consciousness: conscious processes « emerge » from neural processes, the latter « entail » of « give rise to » conscious properties by the « phenomenal transformation » that results in qualia; qualia « reflect » neural differentiations. The view behind Edelman's theory is that consciousness is a process, whose structural properties can be traced back to structural properties of its material substratum. Since these words immediately evoke emergentism and epiphenomenalism, one is tempted to ask which ontological framework best fits the theory.

[13] See Dennett [9], pp. 84–95 and Edelman's account on his own 'Darwin' robots in Edelman [15], pp. 125–141.

Layout: T1 Standard SC | Book ID: 305843_1_En | Book ISBN: 978-3-642-37427-2
Chapter No.: 9 | Date: 20-4-2013 | Page: 11/16

5 Phenomenology and Nature: Edelman's Theory and the Problems of Contemporary Science of Consciousness

Edelman's theory of consciousness, with its deep intertwining of mechanistic models and phenomenology, presents an interesting case for contemporary science of consciousness and philosophy of mind. Though formulated as a naturalistic completion of Darwin's program, it is coherent with some anti-naturalistic claims made in contemporary phenomenological approaches. This suggests a number of methodological and ontological remarks on contemporary research.

First, Edelman's theory asserts the heuristic primacy of consciousness over mechanical explanation: the phenomenological characterization of consciousness by means of introspection is a preliminary stage of modeling and empirical research. The same conclusion has been supported in the phenomenological tradition and receives a growing attention among both philosophers and neuroscientists.[14] But agreement with this simple observation does not imply any "phenomenological" turn. The point, here, is simply that one must focus the explanandum before providing the explanation. This is even true of Dennett's theory, which aims at explaining away all the phenomenological contents, and in order to do so starts with a third-person description of subjective experience, which Dennett calls "heterophenomenology": phenomenology, therefore, must be on stage in order to be criticized. In this sense, Edelman's theory clearly shows that there is no substantial methodological contrast between naturalistic and phenomenological approaches.

Compared to the phenomenological perspective, on the other hand, the evolutionary background of Edelman's ideas allows of a different way to bridge the gap between natural science and humanities. The metaphorical and creative character of human thought adequately reflects the multidimensional system of neural activity, whose interaction with the environment is subject to the mechanism of « recognition of configuration » : insofar the description of neural networks is able to catch the overall features of experience. Nonetheless, with the development of language and culture, the properties of the « second nature » produce an autonomous domain of sense, where scientific hypotheses themselves—including neural models—are constructed, and which cannot be in itself ever subject to neurological description:

> « Although it is true that a scientific description of the world hews more closely to the structure of that world than do our daily impressions, our account of how the brain works suggests that scientific hypotheses themselves emerge from ambiguous (and occasionally irreducible) properties that give rise to pattern recognition. The brain structures and

[14] An agreement on this point was already reached by Jean-Pierre Changeux and Paul Ricoeur in their dialogue on the neurology of consciousness [6]. For a first introduction on phenomenological methods see Gallagher/Zahavi [17].

dynamics leading to such properties are scientifically describable, even if the properties themselves cannot be fully reduced » ([15], p. 146).

This leads to the more complicated issue of ontology. By stressing the existence of qualia in a naturalistic framework Edelman provides an internal critique of reductionism, that apparently agrees with some points made by phenomenologists about the impossibility of modern science to catch the « world of life ». Indeed Edelman's theory has been appreciated by Searle, which is one of the most strenuous supporter of the impossibility to reduce phenomenology to material properties. Searle agrees with Edelman's statement that to assert the neural origin of conscious properties does not imply the latter's « reduction » to neural structures and elaborating on this point he presents an argument against Dennett's denial of qualia.[15] Nonetheless—striking as these analogies may appear—Edelman's theory presents a slightly different perspective.

The difference with Searle is signaled by a basic disagreement: Edelman considers his own model as a satisfactory scientific sample of a theory of consciousness, whereas Searle, as we have seen, considers it unsatisfactory and unable to solve the « mystery » of consciousness. But in order to understand the originality of Edelman's perspective it is useful to consider it in the context of a crucial problem of contemporary philosophy of neuroscience: that is, how to derive consciousness from a multiplicity of physical elements. Dennett himself has traced this problem back to Leibniz's claim that unity of perception cannot be caused by mechanical processes, claiming that Leibniz conflated the epistemic problem of giving this sort of explanation (which was not possible at Leibniz' times and—on the contrary—would be possible in contemporary A.I.) with an ontological verdict ([11], pp. 3–7). Dennett himself considers the subdivision of the « personal level » to « subpersonal » levels as a fundamental heuristic move. The idea behind Dennett's « intentional stance » —the attribution of conscious properties, such as belief and desire, to machines or to parts of the brain—is that

> « when we engineer a complex system (or reverse engineer a biological system like a person or a person's brain) we can make progress by breaking down the whole wonderful person into subpersons of sorts agentlike systems that have part of the prowess of a person, and then these homunculi can be broken down further into still simpler, less personlike agents, and so forth—a finite, not infinite, regress that bottoms out when we reach agents so stupid that they can be replaced by a machine » ([3], p. 88).

Now, without entering the details of Dennett's hypothesis, the claim made here is that we can conjecture a theory where physical parts compose a conscious whole, without properly attributing intentionality, or any other conscious property, to the parts themselves. As Dennett puts it, « we don't attribute fully fledged belief

[15] « Dennett denies the existence of the data to start with. But couldn't we disprove the existence of these data by proving that they are only illusions? No, you can't disprove the existence of conscious experiences by proving that they are only an appearance disguising the underlying reality, because *where consciousness is concerned the existence of the appearance is the reality* » ([26], 112).

Layout: T1 Standard SC | Book ID: 305843_1_En | Book ISBN: 978-3-642-37427-2
Chapter No.: 9 | Date: 20-4-2013 | Page: 13/16

(or decision or desire—or pain, heaven knows) to the brain parts […] No, we attribute an attenuated sort of belief and desire to these parts » ([3], p. 87). It is not entirely clear whether—and in which sense—any of these parts can be considered as a real intentional entity; Dennett himself considers the issue unimportant, since « the security of our intentional attributions at the highest levels does not depend on our identifying a lowest level of real intentionality ». This implies, according to Dennett, that we do not need to give up our standard concept of matter and can rely on the underlying scientific theories.

By making this claim Dennett is perfectly aware of an alternative way, which is to look for new physical (or psycho-physical) theory and, consequently, to modify our basic physical concept of matter. This bold move is made in a number of different ways in contemporary research: without going back to Eccles' « psychons », one can think of Penrose's « microtubules » hypothesis and of Chalmers' reference to a still undeveloped physical theory where consciousness would be a fundamental property of nature. Dennett considers all this as mere speculation (see e.g.: Dennett [9], pp. 36–37; Dennett [11], pp. 8–10), and yet this kind of speculation has found some support by one of the leading researchers in the neurology of consciousness.

Antonio Damasio considers a monistic ontology as the only reasonable background for a solution of the mind–body problem. In order to find a philosophical framework to this conviction Damasio has positively reconsidered monistic metaphysics of the past, drawing from ideas of Spinoza and Whitehead for his theory of the self.[16] Even though he does not accept any of these metaphysical theories as such, Damasio—breaking with his own dualistic terminology of the past—in his last book *Self comes to Mind* presents extension as an attribute of the mind and terminologically identifies 'neural patterns', 'maps' and 'images' ([8], p. 15, 64). This is a significant step, compared to the previous recognition of a « isomorphism » between images, neural patterns and objects ([7], p. 200). Now isomorphism is considered as a sign of objective identity. In a footnote Damasio cautiously questions the « traditional conceptions of matter and mental » as « unnecessarily narrow » ([8], p. 322, n. 14). Though recognizing that « the burden of proof does rest with those who find it natural for mind states to be constituted by brain activity », Damasio does not hesitate to set out his hypothesis: the « looped circuit » of signals transmitted between body and nervous system would enact a « functional fusion of body states and perceptual states » and—by going still deeper into the neuron circuit level—it would be possible to attribute a « protocognition » and « protofeeling » to single neurons, whose joint activity creates the conscious mind.[17]

[16] See e.g. Damasio [7], pp. 184–220, 308n.

[17] Damasio [8], pp. 256–258. While « protocognition » would correspond to the synchronic activity of a nested hierarchy of neurons (p. 252), « protofeeling » would depend on the inherent « sensitivity » or « irritability » of single cells, itself corresponding to the ability to detect and respond to changes inside and outside the cell membrane, that simple organisms display in order to preserve the homeostasis and protect the integrity of the living tissue (pp. 31–60; 258).

Layout: T1 Standard SC | Book ID: 305843_1_En | Book ISBN: 978-3-642-37427-2
Chapter No.: 9 | Date: 20-4-2013 | Page: 14/16

Here we find an interesting (and surprising) analogy with Dennett: the parts of the nervous system are endowed with a "protocognition" and "protofeeling". A huge difference lies of course in the completely different scientific framework of the two hypotheses: so Damasio really regards his line of inquiry about the properties of organic matter as « worth pursuing » , whereas Dennett relies on artificial network models and does not engage the speculative issue of the "consciousness" of cells.

These different speculations may be considered alternatively as a sign of advancement or disorientation, but there is anyway an evident problem that they all must face. That is, since we still do not have a fully articulated and successful theory of consciousness as grounded on subsytems (whether biological or artificial), both possibilities are logically open: to change the scientific theory or to change the basic concepts. As long as the hypotheses cannot rely on decisive empirical evidence, the conflict between Dennett's "multidraft" model (grounded on engineering and A.I. models) or Damasio's self theory (grounded on anatomical and biological data)—and, for that matter, Penrose's speculations on quantum–mechanical foundations of mind—cannot be settled in any definitive way.[18] Even though a number of plausibility claims can be made, and even though—for example—Damasio relies on some more factual evidence than Dennett, there is no way to decide which kind of hypothesis will lead towards a full-fledged theory of consciousness, as being grounded in physical parts of the organism. This doubles the uncertainty, as the methodological doubt is connected to a doubt about the ontological (or metaphysical) background of scientific description.

In this fragile and open context, lest we do not dismiss the whole contemporary research as « frustrating » because the subject is plagued with old mistakes ([26], p. xi),[19] and wait for a solution to be found by ways of physical or metaphysical

[18] Penrose's speculation involves the interpretation of some problematic features of Quantum mechanics, and therefore seems to add problems to problems. In quantum mechanics itself there is a similar (but, in a sense, reversed) epistemological problem: the standard theory includes a problematic interaction between observer and physical system, which has offered space for speculation and criticism; on the other hand, alternative theories (such as Bohmian Mechanics and Collapse models) are not supported by better empirical evidence and involve different conceptual and mathematical problems. For an overview see Pecere [24].

[19] Commenting on Damasio's last book, Searle criticizes his distinction between mind and consciousness, and denies that Damasio's book presents any advancement towards the solution of the « mystery of consciousness » [27]. I think that Searle, sticking to the « standard understanding of the causal relation between mind and brain », misses Damasio's point, which is to radically object to this standard view and elaborate a monistic ontology, where mind and neural patterns are two aspects of the same process which underlies consciousness. This does not mean that Damasio's theory of consciousness, which I cannot examine here in detail, is complete and free of argumentative problems (indeed, it is not). But I think that Searle's reasons of dissatisfaction lead too hastily to the usual conclusion, repeated in reply to a reader of the quoted review: « the way neurons produce consciousness remains mysterious » and « we may never have a solution to the mystery of consciousness ».

Layout: T1 Standard SC | Book ID: 305843_1_En | Book ISBN: 978-3-642-37427-2
Chapter No.: 9 | Date: 20-4-2013 | Page: 15/16

speculation (with the risk of granting the views of those who challenge the validity of standard scientific inquiry), Edelman's theory appears to offer a provisionary, reasonable standpoint, developed along the classic pathway of post-Newtonian methodology of natural science. He draws a parallel between his own experimentally provable correlation of conscious states (C) with neural states (C') and the proportionality set by the formula: F = mA ([14], p. 146). Trying to set up a measurement system, Edelman considers consciousness as a matter of experimental evidence and a property which comes in different degrees. It can be observed that, in the classical Newtonian framework, this means not to address consciousness as an essential property, while leaving open to successive inquiries whether it can have a further explanation. Nevertheless consciousness appears as a true property of organic matter, which still awaits a better understanding, but which—being measurable and possibly subject to a lawlike description—is not in itself a mystery. Here is Edelman's "Newtonian" reply to the charge of not having explained the "actual feeling of a quale":

> « these are the properties of the phenotype, and any phenotype that is conscious experiences its own differential qualia because those qualia are the distinctions made. It suffices to explain the bases of these distinctions—just as it suffices in physics to give an account of matter and energy, not why there is something rather than nothing » ([14], p. 146).

These claims can be made without committing to a particular ontological framework, such as epiphenomenalism (notwithstanding some evident similarities to this approach which are undeniable in Edelman's writings, such as his denial of conscious causality[20]).

We get therefore to some philosophical conclusions that all the quoted antagonists in contemporary research may share, since they do not require to completely settle neither the scientific, nor the ontological issues that we have discussed: consciousness is no mystery, although the description of its neural correlates is still in its early development (and could modify by way of this development our scientific or ontological tenets); but even when possessing such a description, we would not have dismissed the phenomenal and linguistic reality as a fundamental and irreducible feature of our experience. Thereby Edelman's theory is able to provide—better than Dennett's controversial and ontologically more committed account—a naturalistic background to contemporary research on the neural correlates of consciousness.

[20] For Edelman's own reply to the « charge » of epiphenomenalism see Edelman [14], pp. 81–85, 145. Edelman denies causal interaction between consciousness and brain processes, being consciousness « entailed » by these very processes. On the other hand, he denies that we are automata, because of the variability of consciousness as a reflection of the complex interaction of the plastic brain with the environment. Moreover he does not deny the role of secondary (language based) consciousness in long term planning.

Layout: T1 Standard SC | Book ID: 305843_1_En | Book ISBN: 978-3-642-37427-2
Chapter No.: 9 | Date: 20-4-2013 | Page: 16/16

References

1. Adam, C., Tannery, P. (eds.): Oeuvres de Descartes. Vrin, Paris (1964–1974)
2. Bennett, P., Hacker, P.M.S.: Philosophical Foundations of Neuroscience. Blackwell, Malden (2003)
3. Bennett, M., Dennett, D., Hacker, P., Searle, J.: Neuroscience and Philosophy. Columbia University Press, New York (2007)
4. Brook, A., Ross, D. (eds.): Daniel Dennett. Cambridge University Press, Cambridge (2002)
5. Changeux, J.-P.: L'homme neuronale. Fayard, Paris (1983)
6. Changeux, J.-P., Ricoeur, P.: Ce qui nous fait penser. La nature et la régle. Odile Jacob, Paris (1996)
7. Damasio, A.: Looking for Spinoza. Vintage Books, London (2004)
8. Damasio, A.: Self Comes to Mind. Constructing the Conscious Brain. Pantheon Books, New York (2010)
9. Dennett, D.: Consciousness Explained. Little, Brown & Co., Boston (1991)
10. Dennett, D.C.: With a little help from my friends. In: Ross, D., Brook, A., Thompson, D. (eds.) Dennett's Philosophy. A Comprehensive Assessment, pp. 327–388. MIT Press, Cambridge (2000)
11. Dennett, D.C.: Sweet Dreams. Philosophical Obstacles to a Science of Consciousness, MIT Press, Cambridge (2005)
12. Edelman, G.M.: Remembered Present. A Biological Theory of Consciousness. Basic Books, New York (1989)
13. Edelman, G.M.: Bright Air, Brilliant Fire: On the Matter of the Mind. Basic Books, New York (1992)
14. Edelman, G.M.: Wider than the Sky. The Phenomenal Gift of Consciousness. Yale University Press, Yale (2004)
15. Edelman, G.M.: Second Nature. Brain Science and Human Knowledge. Yale University Press, New Haven (2006)
16. Edelman, G.M., Tononi, G.: A Universe of Consciousness. How Matter Becomes Imagination. Basic Books, New York (2000)
17. Gallagher, S., Zahavi, D.: The Phenomenological Mind. Routledge, Abingdon (Oxon)/New York (2008)
18. Garber, D.: Understanding interaction. What Descartes should have told Elisabeth. In: Id., Descartes Embodied. Reading Cartesian Philosophy through Cartesian Science. Cambridge University Press, Cambridge (2001)
19. Leibniz, G.W.F.: Monadology. In: Gerhardt, C.J. (hrsg.), Die philosophischen Schriften, vol. VI, pp. 607–623. Lorentz, Leipzig (1932)
20. McLaughlin, B.P.: The Oxford Handbook of Philosophy of Mind, Oxford University Press, Oxford (2009)
21. Moravia, S.: The Enigma of the Mind. The Mind-Body Problem in Contemporary Thought. Cambridge University Press, Cambridge (1995)
22. Stenon, N.: Discours sur l'anatomie du cerveau. Robert de Ninville, Paris (1669)
23. Tononi, G.: An information integration theory of consciousness. BMC Neurosci. **5**, 42 (2004). http://www.biomedcentral.com/1471-2202/5/42
24. Pecere, P.: Fisica quantistica e realtà. Considerazioni storico-filosofiche. In: Argentieri, N., Bassi, A., Pecere, P., Meccanica quantistica rappresentazione realtà. Un dialogo tra fisica e filosofia. Bibliopolis, Napoli (2012)
25. Schneider, S.: Daniel Dennett on the nature of consciousness. In: Velmans/Schneider (eds.): The Blackwell Companion to Consciousness, Blackwell, Malden, pp. 313–324 (2007)
26. Searle, J.: The Mystery of Consciousness. The New York Book Review, New York (1997)
27. Searle, J.: The Mystery of Consciousness Continues. In: « The New York Review of Books » , June 11, 2011, http://www.nybooks.com/articles/archives/2011/jun/09/mystery-consciousness-continues/?pagination=false#fnr-2

Layout: **T1 Standard SC** Book ID: **305843_1_En** Book ISBN: **978-3-642-37427-2**
Chapter No.: **9** Date: **20-4-2013** Page: **17/16**

28. Velmans, M., Schneider, S. (eds.): The Blackwell Companion to Consciousness, Blackwell, Malden (2007)
29. Voss, S.: Descartes: Heart and Soul. In: Wright, J.P., Potter, P. (eds.) Psyche and Soma. Physicians and Metaphysicians on the Mind-Body Problem from Antiquity to Enlightenment, Oxford University Press, Oxford (2000)