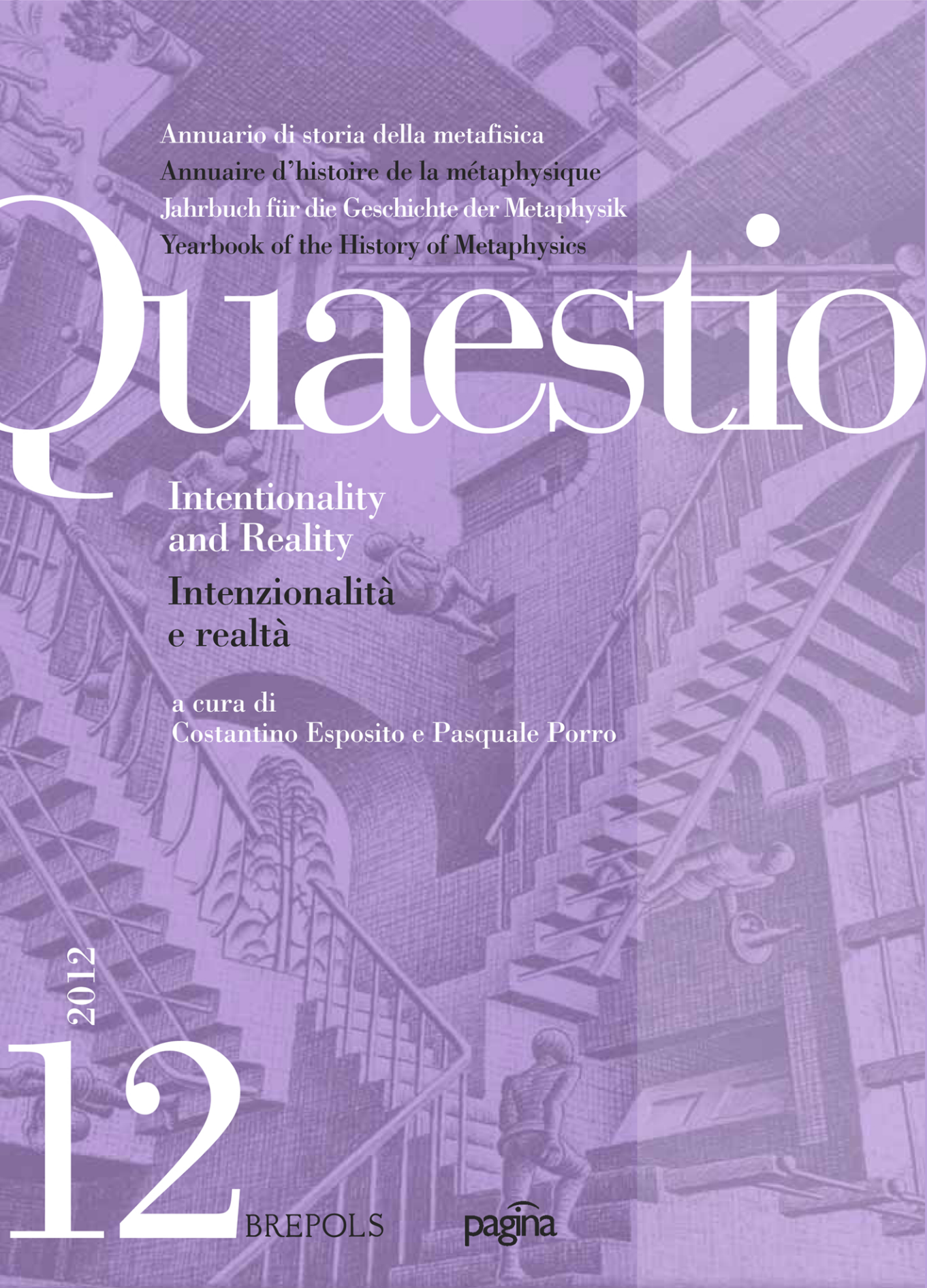

Annuario di storia della metafisica
Annuaire d'histoire de la métaphysique
Jahrbuch für die Geschichte der Metaphysik
Yearbook of the History of Metaphysics

Quaestio

Intentionality and Reality
Intenzionalità e realtà

a cura di
Costantino Esposito e Pasquale Porro

12
2012
BREPOLS
pagina

Quaestio 12/2012

Annuario di storia della metafisica
Annuaire d'histoire de la métaphysique
Jahrbuch für die Geschichte der Metaphysik
Yearbook of the History of Metaphysics

Questo numero di *Quaestio* è stato realizzato anche grazie a un contributo del Consiglio di Amministrazione dell'Università degli Studi di Bari Aldo Moro.

ISBN 978-2-503-54355-0
ISSN 1379-2547
D/2013/0095/238

Quaestio 12/2012

Intentionality and Reality

Intenzionalità e realtà

a cura di Costantino Esposito e Pasquale Porro

BREPOLS

pagina

Quaestio is a peer-reviewed journal, open to unsolicited contributions.
The articles sent to the Editors are normally assessed by a member of the Advisory Board and another specialist chosen by the Board, or by two external specialists. The Editors will maintain records of the reviewers, though their identity will not be made public.

Contributi e volumi per recensione vanno inviati alla Direzione di «Quaestio» (Please send contributions and review-copies to):

Costantino Esposito • Pasquale Porro
Dipartimento di Filosofia, Letteratura, Storia e Scienze Sociali (FLESS) -
Università degli Studi di Bari Aldo Moro
Palazzo Ateneo - Piazza Umberto I, I-70121 - Bari (Italia)
e-mail: costantino.esposito@uniba.it • pasquale.porro@uniba.it

Abbonamenti / Abonnements / Subscriptions
Brepols Publishers, Begijnhof 67 - B-2300 Turnhout (Belgium),
tel. +32 14 44 80 20 - fax +32 14 42 89 19
e-mail: info.publishers@brepols.com

Indice/Contents

Costantino Esposito / Pasquale Porro
Premessa IX

Intentionality and Reality

Jörg Alejandro Tellkamp
Francisco Suárez on the Intentional Species 3

Francesco Marrone
Ontologia dei contenuti ideali.
Essere oggettivo e realtà nel dibattito Descartes-Caterus 25

Giovanna D'Aniello
L'*intentio animi* nella sintesi agostiniana di Cornelis Jansen 79

Martin Lenz
Locke's Theory of Ideas and the Myth of the Given 101

Alain de Libera
L'Ouverture écossaise : Brentano critique de Bain 123

Donatella Colantuono
Conoscenza della realtà e realtà come conoscenza.
Il punto di vista di Bernard Bolzano 153

Laurent Cesalli / Hamid Taieb
The Road to *ideelle Verähnlichung*. Anton Marty's Conception
of Intentionality in the Light of its Brentanian Background 171

Mauro Antonelli
Thoughts Concerning Anton Marty's Early Conception of Intentionality.
Was He Thinking what Brentano Was Thinking? 233

Pierre-Jean Renaudie
Le réel et ses signes : Brentano et Husserl
sur l'engagement ontologique de l'intentionnalité 243

Carmine Di Martino
Husserl: intenzionalità *versus* realismo? 263

Wouter Goris
Das historische Apriori bei Husserl und Foucault – Zur philosophischen
Relevanz eines Leitbegriffs der historischen Epistemologie 291

Miguel García-Valdecasas
Wittgenstein on Intentionality and Representation 343

Oscar Barroso Fernández
Actualidad e intencionalidad en Zubiri 369

Elisabetta Sacchi / Alberto Voltolini
To Think Is to Literally Have Something in One's Thought 395

Stefania Scardicchio
Idealism, New-Realism and Pragmatism:
The American Debate on Reality from Royce to Lewis 423

Paolo Pecere
Naturalizing Intentionality between Philosophy and Brain Science.
A Survey of Methodological and Metaphysical Issues (1969-2011) 449

New Realism: A Debate

Pasquale Porro
A Few Questions about the New Realism 484

Maurizio Ferraris
Responses to *Quaestio* 485

Mario De Caro
The Two Faces of Realism 503

Costantino Esposito
Truth without Certainty? An Open Question into New Realism 515

Matteo Morganti
New Realism and Scientific Realism 535

Andrea Sereni
Mathematical Realism: What's New? 551

Varia

Marienza Benedetto
Da Atene a Baghdad (e oltre).
Ancora sul rapporto tra filosofia araba ed eredità greca 569

Bernd Goehring
Henry of Ghent on Human Knowledge and Its Limits 589

Note Cronache Recensioni

Robert Theis
Transzendentalphilosophie zwischen Scholastik und Kritik 617

Sabino Paparella
Gli "altri" Levinas: ontologia, tempo e politica
a colloquio con la *pensée nouvelle* 620

Indice dei nomi 629

Paolo Pecere

Naturalizing Intentionality between Philosophy and Brain Science

A Survey of Methodological and Metaphysical Issues (1969-2011)

1. Philosophical naturalism, intentionality and brain science: introductory remarks

To give an account of intentionality in terms of the concepts and methods of natural science has been a crucial objective of early XXth century "scientific philosophy". This objective has been originally motivated by the neoempiristic project to reduce the intentional idiom to a physicalistic language, with the aim of ruling out metaphysical and unverifiable propositions about mental phenomena. Following Sellars' (1952), Chisholm's (1957) and Quine's (1960) contestation of the possibility of reducing the intentional idiom to a nonintentional language, the crucial task of denying Brentano's thesis, according to which no physical phenomena exhibits intentionality, has been pursued by naturalistic philosophers and cognitive scientists in a number of different ways. Generally speaking, the attempt at naturalizing intentionality in cognitive science has often represented a privileged way to argue for a naturalization of mental phenomena in general, thereby supplementing or substituting straightforward materialistic claims of the past, because intentionality – contrary to consciousness or qualitative feeling – has appeared reducible to syntactic or biological structures[1]. Eliminationists and materialists (such as Paul and Patricia Churchland and Daniel Dennett) have argued that the intentional idiom fails to describe any real property, and intentionality can be "explained away" by means of computational accounts of neural networks. Others have argued that intentionality is indeed a real property, albeit not an exclusive one pertaining to mental phenomena, as it also belongs to sentences of natural languages and other biological phenomena (e.g. Dretske, Millikan, Tye). On the other hand, advocates of the reality *and* uniqueness of conscious intentionality (e.g. Searle and Chalmers) have claimed that it can be

[1] For a survey see P. Jacob, *Intentionality*, § 6, in E.N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2010 Edition), URL = <http://plato.stanford.edu/archives/fall2010/entries/intentionality/>.

nevertheless considered as a natural property, without the need to deny its original first-person phenomenology, with the consequence that standard natural science itself has to be accommodated to the treatment of this phenomenon.

A common feature of these different naturalistic approaches has been the need to define the content and role of standard natural sciences, and the possibility of naturalization has been often accompanied by a certain anticipation of future empirical results or a speculation about a possible future modification of experience itself as a consequence of scientific knowledge[2]. Gradually, the need for a critical evaluation and interpretation of given scientific theories by means of independent epistemological and ontological arguments has become evident among both philosophers and neuroscientists, and this process has often involved the reappraisal of classic metaphysical frameworks. I will hence consider this superposition of epistemology and metaphysics in a number of case studies, focusing on the peculiar interaction between philosophy and neuroscience.

2. Critique of phenomenological evidence and evolutionary theory: Dennett and Millikan

The works of Daniel Dennett, which constitute a landmark for contemporary discussions, are driven by a strong conviction of the sufficiency and fruitfulness of standard natural science for a reductive account of intentionality. In *Content and Consciousness* (1969) Dennett sets out to connect the common sense views about mental events to biology and artificial intelligence: "What is the relation between a man's mental life and the events in his brain? How are our commonplace observations about thinking, believing, seeing, feeling pain to be mapped on to the discoveries of cybernetics or neurophysiology?"[3]. According to Dennett, if we approach these questions by admitting the existence of both minds and bodies, then we are led to the post-Cartesian "fruitless pendulum swing" of metaphysical hypotheses, such as interactionism, monism and parallelism (p. 3). In order to introduce a new way out of this problem, Dennett confronts the two main

[2] This is already evident in the history of identity theory and materialism of the '30s and '50s, from Feigl to Rorty: identity of mental and physical phenomena was a metaphysical hypothesis, recommended by the empirical success of physical and biological sciences, which the future development of these sciences itself would eventually confirm, leading to a necessary modification of standard vocabulary about mental states. Subsequently, Popper would criticize this view, arguing that philosophers present a "promissory note" in place of a monistic scientific theory. See S. Moravia, *L'enigma della mente. Il 'mind-body problem' nel pensiero contemporaneo*, Laterza, Roma-Bari 1986, pp. 20-21, 27, 104-109 (engl. transl.: Id, *The Enigma of the Mind. The Mind-Body Problem in Contemporary Thought*, Cambridge University Press, Cambridge 1995).

[3] D. Dennett, *Content and Consciousness*, Routledge, London 1969 (1986[2]), p. 13. Page numbers given in brackets will always be referred to the book quoted in the last footnote.

contemporary attempts of connecting intentionality to natural science, behaviourism and phenomenology, arguing that none of the aforesaid provide for satisfactory solutions: in fact, the former did not succeed in explaining the learning and goal-directedness of behaviour itself, while the latter provides intentional psychological explanations that are in strong contrast with the rest of science, since they do not account for the complex biological conditions of animal behaviour and do not require any such account (pp. 32-39). The way out is the ascription of intentional expressions to internal physical events (Dennett will refer to this operation as the "subpersonal" level of explanation) (p. 39). By confirming a similar hypothesis, we will be justified in ruling out the existence of pains and desires from the scientific vocabulary and therefore reduce intentionality to a matter of description with no ontological value. Dennett's claim in this book is that there are "promising hints" (p. 40) suggesting a positive completion of this program with the cooperation of computer science, evolutionary theory and brain science.

Dennett has later found a substantive foundation of his ideas on language and thought in the work of Ruth Millikan. Millikan elaborates a full-fledged theory of language and thought, both considered as forms of intentionality, which heavily relies on evolutionary biological arguments. Millikan's negative objective is to challenge what she calls "Meaning Rationalism", that is the widespread thesis that we can grasp the meaning of propositions by introspection, independently of the consideration of their function in linguistic communication:

> "we must be willing to discover that, just as we cannot know a priori or with Cartesian certainty whether any particular thing we think or say is true, so we cannot know a priori or with Cartesian certainty that in seeming to think or talk about something we *are* thinking or talking about – *anything at all*. We cannot know a priori *that we mean*"[4].

According to Millikan intentionality is a property of "language devices" which depends on "proper functions", that is, "functions that explain the survival or proliferation of these devices together with their characteristic cooperative hearer responses" (p. 5). These functions, as the functions of any biological device (including organs), can only be defined by referring to "reproductively established families" of tokens, that is to the "history" of these devices in populations of language users, for it is this history that shows what the devices are "supposed

[4] R.G. Millikan, *Language, Thought and Other Biological Categories*, The MIT Press, Cambridge Mass. 1984, p. 10 (the book includes a sympathetic Foreword by Dennett). "Meaning Rationalists" hold that "intension can't be wrong or mistaken and that mere (seeming) thoughts-of, as opposed to judgments about, cannot be senseless". They include Descartes, Hume, Wittgenstein (in the *Tractatus*), Quine and even an externalist and realist such as Putnam (pp. 325-326 and 328-333 for the critique of Putnam).

to" mean, and the meaning of a physically identical device (whether organ or sign) can change in time. Independently of this context, it makes no sense to ascribe a function to a device – both natural and artificial (see pp. 17-38). By applying this approach to mental representations, meaning appears as a relative content, which cannot be fixed by introspection, and intentionality as "of-ness" of "aboutness" is just a "by-product of the notion of 'proper function' " (p. 5), which does not require for its definition any "given" mental representation.

As Millikan makes clear in a later restatement of his arguments, this teleological reasoning does not provide by itself a full theory of representation, but provides a constraint for any such theory[5]. Indeed, in *Language, Thought and Other Biological Categories*, Millikan does not deny (on the ontological level) that there *are* intentional representations "in [man's] thought", but only argues that they are not "transparent", that is, they require an interpretation of their function in a biological and linguistic context in order for their meaning to be fixed (p. 13). Representations are "relations between the head and the world" (p. 331), and therefore mistaken representations are just empty and possess no "intentional object"; they are "failing" representations, but still they exist in the mind.

Dennett draws on Millikan for the refinement of his different, instrumentalist conception of intentionality and mind presented in *The Intentional Stance* (1987)[6]. A major feature of contemporary theories of intentionality in cognitive science is the distinction between an "original" or "real intentionality", typical of mental states, and "derivative" intentionality, which signs and artifacts derive from the intentions of men. Dennett's general objective consists in challenging this widespread distinction, which is typical of "meaning rationalists" (in Millikan's sense), arguing that there is nothing as a *real* intentionality and an ultimate mental source of meaning. Dennett's main argument is that the intentional meaning of human expressions and behaviour only depends on the "proper function" in Millikan's sense, and therefore the real source of intentionality is "Mother Nature" (pp. 298-300). In a variant of a typical mental experiment of contemporary theories of meaning, Dennett imagines a man on a Twin Earth where "schmorses" exist in the place of perceptually indistinguishable "horses", and these animals are called "horses" by local speakers. In this case, so Dennett ar-

[5] R.G. Millikan, *Varieties of Meaning*, The MIT Press, Cambridge Mass. 2004, pp. 63-67.

[6] See D. Dennett, *Evolution, Error and Intentionality*, in Id., *The Intentional Stance*, The MIT Press, Cambridge Mass. 1987, pp. 287-321 (successive references are given in brackets). This chapter includes several positive references to Millikan and interesting critical discussions of the views on intentionality developed in this period by Searle, Fodor, Dretske, Burge, Putnam and Kripke, which are all considered as "Meaning Rationalists".

gues, there would be no way to tell – and it would indeed make *no sense* to ask – whether the speaker *really* means one or the other meaning (pp. 294-295).

Dennett inserts Millikan's critique of meaning rationalism (including phenomenology) into his own materialistic explanation of consciousness, adding the additional (and ontological) thesis that there are no inner representations in the brain. This hypothesis is further articulated in *Consciousness explained* (1991). The overall view advanced by Dennett is that there is no way to empirically establish at which point, in the chain of spatio-temporal events that connect environmental stimuli to an organism's behaviour, a conscious representation takes place and is given to a subject (an analogous claim can be made regarding non-intentional conscious states such as pain). Dennett calls this hypothetical place "Cartesian Theater" and the mistaken theory that posits the existence of this place "Cartesian Materialism". In Dennett's view, phenomenological evidence of "given" perceptions and thoughts is just a "belief", which can be explained by means of analysis of the "mechanisms" that underlie behaviour and verbal production, according to a "Multi-Draft" model which does not require a central coordinating function to be located in the brain. In the light of his theory, Dennett considers *qualia* and the self as abstract objects, "fictions" of the same kind as "centers of gravity"[7].

Without entering Dennett's complex and controversial theory into further detail, we can now consider some methodological aspects of his approach to intentionality. Dennett considers philosophy and science as "continuous" enterprises: philosophy explores possibilities and examines hypotheses, by making large use of mental experiments, while science's role is to provide theoretical details and empirical confirmation[8]. Wittgenstein's philosophy of psychology provides an exemplary method to critically eliminate illusory issues in the light of a pragmatic analysis of the language (the critique of "private language" in the *Philosophische Untersuchungen* is a clear antecedent to Dennett's critique of meaning and feeling). However, because Wittgenstein's standpoint rigidly separates philosophical and empirical claims, Dennett considers it just a station towards a scientific theory of the subpersonal correlates of intentionality in the brain and, in order to elaborate this theory (and its justification), he refers to different scientific fields of research. From this point of view, evolution theory is a prime source of arguments, since it provides the model for functional explanations with no intelligent design, which Dennett extends by analogy to the philos-

[7] D. Dennett, *Consciousness Explained*, Back Bay Books, New York 1991, pp. 102-138 (in part. pp. 134-135 for a summary of the theory). On the status of *qualia* and the self (which Dennett calls "heterophenomenological objects") see pp. 95, 96-98. On *qualia* see also pp. 321-411, on the self pp. 412-430.

[8] See Dennett, *The Intentional Stance* cit., p. 5.

ophy of language and mind. But this is still not enough to explain the functioning of the human brain with its peculiar differences as compared to abstract computing systems. In order to instantiate his theory, an evolutionary mechanism of selection of mental functions must be conceived and detected in the brain.

In *Content and Consciousness* Dennett sketches an evolutionary hypothesis on the development of neural structures[9]:

> "Given a brain with an initial plasticity or capacity for producing different functional structures as a result of input, the key to utility in the brain must be the further capacity to sort out these functional structures, keeping and using those that are useful to the survival and comfort of the organism and eliminating or refraining from using the harmful ones" (p. 48).

Dennett roughly suggests some hypothesis about how this mechanism of natural selection of functional structures in the brain could work, figuring out that the brain must have stimulating and inhibitory functions. The purpose of this hypothesis, however, is "to reveal the general shape such a theory must have" (p. 71).

In the Preface to the Second edition of *The Intentional Stance* (1985) Dennett observes that contemporary scientists "are now developing 'evolutionary' models at a level of empirical detail and sophistication", but he no longer develops the hypothesis[10]. A short review of contemporary proposals is provided in a footnote of *Consciousness Explained* (p. 184), but in this book Dennett, in order to avoid leaving "too many things mysterious", ascends (in a functionalistic fashion) to "a more general and abstract level" (p. 193) and elaborates a model of consciousness as a "virtual machine implemented in the parallel architecture of the brain" (p. 210).

Still Dennett insists that his philosophical work regards only the "possibility" of a theory, and his "theory sketch or model of how the brain *might* do something" provides a "research program" which eventually must lead to an "empirically confirmed theory" (p. 41). This, in turn, produces a dependence of Dennett's ontological claims on some empirical brain science, which must implement the arguments and hypotheses drawn from computer science, evolutionary biology and linguistics. His main argument for the ontological elimination of "real" intentionality in the mind is consequently heuristic: though he has no "knockdown argument" against dualism he argues that "accepting dualism is giving up" in the task of elaborating an empirical theory of consciousness, by accepting the view that intentionality and consciousness are a "mystery" (p. 37).

[9] See *The Evolution in the Brain* chapter in DENNETT, *Content and Consciousness* cit., pp. 43-71.

[10] DENNETT, *Content and Consciousness* cit., p. XI. See below § 5 for a discussion of the examples quoted by Dennett (including Edelman).

On the whole Dennett's theory, with his ontological denial of *qualia*, ultimately *needs* a mechanistic explanation of syntactic production in neural networks. In the next section we will see how far neuroscientists actually satisfy Dennett's theoretical postulates and whether this agreement can also apply to his conception of intentionality and consciousness.

3. Antireductionist naturalisms: Searle, Chalmers, Putnam

3.1. Searle

Reductionist approaches to intentionality have usually involved the claim that conscious intentional states in human beings (if there are any) are a subset of intentional properties of biological phenomena. A contrary approach has considered human intentionality as coextensive with, or included in consciousness as a genuine mental property. This has led to look for a different way to accommodate intentionality in a naturalistic world view, which depends on the possibility to provide an account of consciousness in terms of natural science.

One of the main advocates of this kind of view has been John Searle. In his book *Intentionality* (1983) Searle argues that "people have mental states which are intrinsically intentional" and that these mental states are necessary in order for any linguistic expression to have a meaning[11]. This view involves a criticism of all other current attempts to solve the "mind-body" problem in a reductive way, by denying the "intrinsic mental phenomena" that accompany any intentional state and considering the latter as purely causal or computational processes (p. 262). This criticism has been spelled out in a number of papers from the early 1980s[12] and finally articulated in the book *The Rediscovery of the the Mind* (1992). It has been directed against behaviourism, functionalism, physicalism, dualism and also to the "intentional stance" strategy, which has been the object of a long controversy between Searle and Dennett[13].

Searle basically defends the reality of "inner subjective qualitative states of

[11] J. Searle, *Intentionality. An Essay in the Philosophy of Mind*, Cambridge University Press, Cambridge 1983, p. viii.

[12] See e.g. J. Searle, *Minds, Brains and Programs* and *Intrinsic Intentionality*, in *Behavioural and Brain Sciences*, 3 (1980), pp. 417-424 and 450-456. Id., *Analytic Philosophy and Mental Phenomena*, in *Midwest Studies in Philosophy*, 5 (1980), pp. 405-423.

[13] J. Searle, *The Rediscovery of the Mind*, The MIT Press, Cambridge Mass. 1992, pp. 5-8 (successive references to this book, in the next paragraph, are given in brackets). For Searle's criticism of Dennett see his review of *Consciousness Explained* in J. Searle, *The Mystery of Consciousness*, The New York Book Review, New York 1997, pp. 97-115, with a short *Exchange* between Dennett and Searle in the *Appendix*, pp. 115-131.

consciousness" and of "intrinsically intentional mental states such as beliefs and desires", and denies the materialistic claim that science has to provide a third-person account of these processes, which can only be described in first-person (pp. XI-XIII). In particular, Searle identifies mental states (including intentional and non-intentional states) with conscious states, considering any other process that takes place in the brain as merely physiological and non-mental, while unconscious states are defined as being theoretically accessible to consciousness (pp. 151-155). With these claims Searle also contrasts the main tenets of a naturalism of the kind defended by functionalists such as Putnam (in the 1970s), Dennett and the connexionists: no computational explanation of mental processes, as grounded in neural networks, can account for their phenomenal content, because computation is "observer relative" and still requires an interpreter. In other words, as Searle' famous mental experiment of the Chinese Room illustrates, no syntactic set of rules can explain the capacity to understand meaning[14].

This antireductionist claim does not intend to deny any scientific, naturalistic explanation of phenomenal consciousness. According to Searle, the real motivation behind behaviourist and functionalist views has been the "fear that unless some way was found to eliminate mental phenomena naively construed, we would be left with dualism and an apparently insoluble mind-body problem". This is also the reason why functionalism (in its different versions) is the "most common contemporary philosophical solution to the problem of intentionality"[15]. The same concern has inspired the program of "naturalizing" intentionality in a third-person description, which is ultimately materialistic. According to Searle an objective account is defined by its being independent from subjective evaluation, and does not logically imply a "third-person ontology". Searle therefore does not accept the claim of "property dualists" (such as Nagel and McGinn) – and, we can add, of Continental phenomenology – which denies the possibility of an account of subjective and intentional states in terms of the standard science of nature. On the contrary, since his first writings on this issue, he argues that consciousness and intentionality are biological properties:

> "On my view mental phenomena are biologically based; they are both caused by the operations of the brain and realized in the structure of the brain. On this view, con-

[14] SEARLE, *The Rediscovery* cit., p. 13. The Chinese Room mental experiment is intended to clarify the distinction between the ability to transform strings of symbols according to a set or rules (which can be implemented in a standard computer) and the mental capacity to understand meaning, which is typical of human brains and depends on their specific causal powers. It is first exposed in ID, *Minds, Brains and Programs*, in *Behavioral and Brain Science*, 3 (1980), pp. 417-457.

[15] J. SEARLE, *Mind. A Brief Introduction*, Oxford University Press, Oxford 2004, p. 162.

> sciousness and Intentionality are as much a part of the human biology as digestion or the circulation of the blood. It is an *objective* fact about the world that it contains certain systems, viz., brains, with *subjective* mental states, and it is a *physical* fact about such systems that they have *mental* features"[16].

This "biological naturalism", while saving the validity of phenomenological, behavioural and physiological data, involves a substantive reference to theories in neurobiology. This connection of philosophy to neurobiology is addressed in the final chapter of *Intentionality*, "Epilogue: Intentionality and the Brain". A crucial point of Searle's biological view of intentionality as "caused by and realized in the brain" is that there are "different levels of description" of the same processes, and this recognition allows to keep together the intrinsically intentional processes of conception and causation with their microbiological description. Searle makes an analogous claim about the solidity of a body, the liquidity of water and the explosion of fuel in an engine, which are also caused by and realised in microscopic processes: it does make sense to consider these properties as causal at the macroscopic level, while this description loses its validity at the microscopic level. The case of mental and intentional states is perfectly analogous: visual experiences and intention in action *cause* behaviour at the macroscopic level, even though at the microlevel the terms "intention in action" and "bodily movement" are "inappropriate" (pp. 267-269). According to Searle, now, "the logical nature of these *kinds* or relations between mind and the brain" is perfectly comprehensible, and thereby the mind-body problem, as it has been historically formulated in post-Cartesian philosophy, can be solved: mental and physical states exist as different levels of description "in the same substance" (Searle quotes the famous mental experiment of Leibniz's *Monadology* about entering into a thinking machine and not being able to find "anything by which to explain perception"). Eventually, it will be perfectly right to localize mental processes in the brain. However, there still is a missing element of this solution: the "empirical and conceptual problems" of describing these relations are "incredibly complex and progress, in spite of much optimistic talk, has been agonizingly slow" (p. 267). For instance, according to Searle, "we know little about how intentional action originates in the brain", but "we do know that neural mechanisms stimulate muscle movements" (pp. 269-270). A similar claim is made, in a recent book, about thirst, and is said to be valid for "more complex" mental processes as well[17].

[16] Searle, *Intentionality* cit., p. ix. The same view has been defended by Searle throughout his successive writings. See e.g. Id., *Mind* cit., pp. 111-115 ("The solution to the mind-body problem").

[17] Searle, *Mind* cit., p. 164.

But at this point a new problem arises, which affects Searle's program until later writings: Searle's main philosophical tenet is to take into account the essential *conscious* aspect of mental processes; but Searle can give no example of the kind of relation that connects the brain to consciousness, as he thinks that no *available* biological theory provides explanations of consciousness, which therefore remains "mysterious". This negative appraisal of contemporary neuroscientific theories casts a shadow on Searle's program of detecting the "causal powers" of the brain[18].

3.2. Chalmers

The objective of realizing a scientific theory of "phenomenal consciousness", or "the way it feels like" to have conscious states, has given rise to a huge literature in the last 20 years (including the foundation of a *Journal of Consciousness Studies* in 1994). The limit of available scientific theories has produced the search for naturalistic accounts in *non-standard* natural science, thereby postulating a reform of science in the name of an ontological claim about mental properties. One of the most discussed books in these studies, *The Conscious Mind* by David Chalmers (1996), pushes antireductionism towards bold metaphysical claims and labels its proposal as "naturalistic dualism"[19]. According to Chalmers, any functional explanation of mental processes or capacities, which can be certainly provided by cognitive science (as the solution to an "easy problem"), does not solve the "hard problem" of consciousness, that is "why is all this processing accompanied by an experienced inner life?" (p. XII). From a logical point of view, Chalmers generally argues that "phenomenal consciousness" is not "supervenient" on physical facts. In order to illustrate his claim he first proposes an old-fashioned "zombie" thought experiment: there could be a physically identical world, where human beings would own all the functional abilities and the same behaviour, while being unable to consciously "feel" as we do (pp. 94-99). This controversial mental experiment is just one among different

[18] Criticism of Crick's, Edelman's and Rosenfield's hypotheses on the neural correlates of consciousness has been set forth in a number of reviews on *The New York Review of Books*, which are included in Searle, *The Mistery of Consciousness*. Dennett's charge that Searle, consequent to his negative appraisal of standard neurobiology, "doesn't have a program of research", is also here, p. 119. A critique of Damasio's theory has been published in a successive review, *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. I will comment on Searle's critique of Edelman in §§ 4-5.

[19] D.J. CHALMERS, *The Conscious Mind. In Search of a Fundamental Theory*, Oxford University Press, Oxford 1996.

counterfactual hypotheses advanced by Chalmers in order to underscore the heterogeneity between physical, biological and computational properties and conscious states, which include intentional states. After anticipating the objections that he may be "setting the standards too high" or that scientific explanation must stop somewhere, Chalmers makes clear that *no available* theory of computational kind (Baars, Dennett, Churchland), in neurobiology (Crick-Koch, Edelman), physics (speculation about quantum physics of Penrose and others) or evolutionary biology successfully accounts for the "hard problem" (pp. 111-122).

Again, by stating this argument Chalmers does not mean to abandon naturalism, but requires a brand new construction of natural science, which has to leave room for a dualistic ontology. Chalmers' proposal is that we need to expand the ontological framework of natural science and look for psycho-physical principles and laws regarding how conscious states are connected to physical states (pp. 213-246). Consequent with his bold challenge, Chalmers does not hesitate to reappraise different metaphysical hypotheses that could provide a framework for a scientific theory of consciousness, including the "panpsychist" hypothesis that "information" itself may have two aspects, a physical and a phenomenal aspect, and therefore even simple processing machines such as a thermostat may have a certain degree of feelings (pp. 276-310).

From a methodological point of view, Chalmers's positive claims are "plausible" hypotheses and "prototheories", that are mainly supported by the negative counterfactual claims about physical reality and phenomenal consciousness: it is not surprising, then, that in the following years Chalmers has declared himself open to many different ways to solve his problem, and much of his work has consisted in the classification and critical evaluation of contemporary proposals within philosophy of mind, much of which have been significantly stimulated by his book. In a recent paper, Chalmers claims that both intentionality and consciousness may represent essential properties of mental phenomena, while insisting that no genuine intentionality can exist without consciousness and reviewing different "possible" theories of representation which could be compatible with the irreducible character of phenomenal properties[20]. Chalmers concludes that his approach on the one hand "offers little prospect for grounding consciousness wholly in intentionality", while on the other hand does not allow to ground "intentionality in consciousness", though "it is not implausible that there is something about consciousness that by its very nature yields a representation of the world"; in the end, intentional content is "part of the essential

[20] D. Chalmers, *The Representational Character of Experience*, in B. Leiter (ed.), *The Future for Philosophy*, Oxford University Press, Oxford 2004, now in Id., *The Character of Consciousness*, Oxford University Press, Oxford 2010, pp. 339-371.

nature of phenomenology that is directed outward at a world" (p. 371). This paper gives one more example of a general methodological aspect of Chalmers' arguments: lacking a full-fledged empirical theory of consciousness, Chalmers heavily relies on the fact that different claims appear to him more or less "plausible", and finally leaves many possibilities open, just putting the constraint that the initial phenomenological evidence must be accounted for in the definitive naturalistic theory.

This result depends on the joint admission of two heuristic principles, "taking consciousness seriously" and "taking science seriously", and by the fact that there is no scientific theory that can provide the ultimate source of validity of any naturalistic "fundamental theory" of the kind theorized by Chalmers. This standpoint of "naturalistic dualism" has been severely criticized by Searle as a sign of "desperation" in cognitive science: it would be a transparent attempt to save both functionalism and the dualistic ontology of property dualism at the cost of uncontrolled and antiscientific metaphysical speculation, while refusing to accept the most obvious way to look for a standard biological account[21]. Nonetheless, Searle expresses dissatisfaction with contemporary scientific accounts of consciousness in a similar way to Chalmers. Indeed Chalmers' insistence on alternative metaphysical hypotheses reveals a typical problem of this kind of naturalistic accounts: while saving phenomenal consciousness, and considering it at the same time as a natural property, these authors cannot ground their claims on existing natural science, and together with this reference they lose a fundamental criterion for the development of a naturalistic world-view, in the anti-metaphysical sense that this approach has taken in the Anglo-American tradition. This produces the inclusion in the space of possibilities of contemporary philosophy of metaphysical hypotheses of the past, from Fechner's psychophysical parallelism to Russell's neutral monism, as well as the production of brand new forms of transcendent hypotheses, which are confronted as having a status of minor or major plausibility and only share the denial of materialism[22].

[21] SEARLE, *The Mystery of Consciousness* cit., pp. 135-163, in part. p. 152 for the charge that Chalmers simply "combines" functionalism and property dualism with "breathtakingly unplausible" consequences. A short exchange between Searle and Chalmers follows at pp. 163-176.

[22] See the review of metaphysical types of hypotheses in D. CHALMERS, *Consciousness and its Place in Nature*, in ID., *The Character of* cit., pp. 103-139 (this chapter is drawn from papers originally published in 2002 and 2003). For a reply to materialistic criticism by Dennett and Paul Churchland see ID., *Afterword: From "Moving Forward on the Problem of Consciousness"* (originally published in 1997), in *The Character of* cit., pp. 28-34.

3.3. Putnam's "liberalized functionalism"

As we have seen, both reductive and anti-reductive naturalistic accounts of intentionality bear an essential reference to neurobiological theories of mind and consciousness. This kind of reference is very common among philosophers working in the field, even by those authors who deny both the possibility to naturalize intentionality and the formulation "hard problem" of consciousness. A remarkable example is Hilary Putnam's recently advanced "liberalized functionalism". Putnam has crossed the whole season of cognitive science, defending a functionalistic view of mental states since the early 1960s, and still maintains that in the light of a functional account of cognitive faculties there is no "hard problem" of consciousness[23]. On the other hand, Putnam has considerably softened his initial position, grounded on the "software-hardware" model for describing the mind-body relation, conceding that "the identification of naturalism with such 'reductions programs' as the program of reducing the intentional to the nonintentional or dispensing with the intensional and normative notions is entirely a mistake". The "liberalized functionalist" still accepts that "functional capacities" are what matter for consciousness and for mental properties, but he (1) does not insist that those functions are internal (thus joining forces with evolutionary accounts in support of the role of the environment), (2) "does not insist that those capacities be described as capacities to *compute*" (thus denying that the "psychological, biological and neurological vocabulary needed to describe those functions [...] be described in a vocabulary drawn from one science (e.g. computer science)" and (3) "does not even eschew intentional idioms", such as "refers to", "if they are needed in describing our functioning, although she naturally wants an account of how intentional capacities grow out of protointentional capacities", thus rejecting a view of naturalizing intentionality that has been widespread in cognitive science. But still this program heavily depends on the work of the scientists: "I see the details as largely something to be worked out by scientists in a number of different fields, but with philosophers playing the necessary, if often unappreciated, role of critics"[24].

Given this shared attribution of argumentative power to brain science by leading philosophers of mind, it is interesting to consider how contemporary scientists in the field have, more or less implicitly, responded to these calls.

[23] H. Putnam, *How to be a Sophisticated "Naïve Realist"* (2011), in M. De Caro / D. Macarthur (eds.), *Philosophy in an Age of Science,* Harvard University Press, Cambridge Mass. 2012, p. 627.

[24] H. Putnam, *Corresponding with Reality* (2011), in Id., *Philosophy in an Age* cit., pp. 82-83.

4. Intentionality in contemporary brain science: three case studies

Standard research in neurosciences, both at the level of basic molecular processes and of complex and distributed neural networks, does not address the issue of explaining general properties such as intentionality (and consciousness). Yet such problems are being considered more and more as frontier fields of research, with the conviction that the rapid development of observational techniques of neural activity, such as EEG, PET and (since the 1990s) fMRI, may lead towards positive empirical results. As a matter of fact, a number of pioneering models of global neural networks and hypotheses on the neural correlates of consciousness have been set out in the last decades[25]. We will consider here three examples of philosophically engaged theories, in order to show how scientists make use of philosophical and metaphysical backgrounds in order to provide an understanding and connection of the different phenomenological, behavioural and physiological data. This shows, in turn, how the very "empirical theories" invoked by philosophers – roughly in the same years – in order to settle the issue of intentionality are themselves grounded on pure theoretical and controversial hypotheses, which contemporary scientists try to force into the methodological framework of empirical research.

4.1. Changeux and materialism

Jean-Pierre Changeux, one of the main contemporary neuroscientists, is the author of a pioneering global workspace model of consciousness and of a number of books about the philosophical meaning of neuroscience, where he defends the need for a physicalistic framework for brain science on the ground of a rich knowledge of the historical and philosophical background of the discipline. In his *L'homme neuronale* (1983) Changeux sketches a critical history of neurological theories, arguing that, after the explanation of long distance neural transmission by means of electrophysiology, the chemical study of neurotransmitters has become the starting point of a full-fledged molecular brain science. Since the 1940s the hypothetical "animal spirits" of early modern physiology have definitively been identified with "atoms and molecules. The sciences of the ner-

[25] According to Kandel the progresses of neuroscience in "the last decade" are beginning to unravel its "deepest riddle", that is the "neural representation of consciousness". E. KANDEL, *The Brain and Behaviour*, in E. KANDEL / J.H. SCHWARTZ /T.M. JESSELL, *Principles of Neural Science*, McGraw-Hill, New York 2000[4], pp. 16-17. For a philosophical appraisal see T. METZINGER (ed.), *Neural Correlates of Consciousness: Empirical and Conceptual Questions*, MIT Press, New York 2000.

vous system have become molecular" and allow to sketch a biological-molecular "explanation" of behaviour[26]. According to Changeux, this progress has meant the recovery of the ancient atomistic hypothesis, which has been long contrasted by metaphysical theories of the soul and can be finally "expressed in complete liberty" (p. 24). Changeux thus regards the whole history of neurology in the light of a straightforward opposition between materialism and spiritualism.

Among the supporting arguments for Changeux's claim is the "spontaneous" *activity* of brain tissue, which is constituted by the electric potential produced by different ionic contributions on both sides of neural membranes and produces communication inside the neural tissue by means of electromagnetic waves (pp. 98, 103). Therefore, Changeux – following the ideas of his teacher, Jean Monod – argues for a neural, *deterministic* explanation of human behaviour by means of the study of the interplay between environmental data and responses (p. 159), and this explanation involves a reductive account of intentionality, which is reduced to the representative power of the brain and does not require any non-material element. This theory, being materialistic, can be said to regard the "brain machine", but it is different from computational theories in AI since it does not set a strong distinction between hardware and software, but rather concerns a complex stratification of topological, electrical and chemical connections, allowing for the "autonomous development" of "strategies" and the "construction of original programs" (p. 161).

On this background, Changeux sketches a "biological theory of mental objects", which are "identified with the physical state produced by the *correlated* and *transitory* activation (both electrical and chemical) of a large population or 'assembly' of neurons distributed at the level of multiple defined cortical areas" (p. 174). By means of electromagnetic observation techniques these objects can be geometrically localized in the neural tissue and their impulses can be expressed in a "graph": thereby a mathematical "model" of behaviour becomes possible (pp. 125-129). By joining this approach to the chemical study of neurotransmitters, it will be possible to provide a completely materialistic account of sensations, as it had been hypothesized in XVIIIth century materialism: " 'Shall the organization be sufficient for everything? Yes, one more time', as La Mettrie wrote. Of course, under the condition of including chemistry!" (p. 146).

Representation of external objects in neural networks requires both originally distinct functions of differently localized neurons and the capacity of the nervous system to develop new patterns of behaviour. The first requirement is nec-

[26] J.-P. Changeux, *L'homme neuronale*, Fayard, Paris 1983, pp. 49, 122-124.

essary in order to account for the most elementary representations, the primary "percepts". Here the isomorphism between the oscillatory graph of the neurons and the object is mediated by the fact that neurons represent sense organs, which, in turn, represent physical features of the environment (pp. 146-152, 176). Sense organs act as "commutators" of "molecular clocks", in that they transform mechanical stimuli into electrical signals, determining the modification of electronic oscillation within neurons according to originally set behavioural functions (p. 107). Given this set of representative functions, which must be considered as the effect of natural selection on the development of the nervous system, Changeux – reappraising typical hypotheses of the modern physiology of the mind – describes the formation of mnemonic images as "traces" of mental objects "distributed over the whole cortex" (p. 203) and concepts as mental objects which partly or totally lack sensory components. The formation of these objects derives from a selective mechanism of "epigenesis", which is inspired by Darwinian theory[27]. In order for a "learning" to take place one has to "postulate" the "pre-representative" activity of the brain, which spontaneously and continuously produces temporary representations (p. 175). The latter are selected by interaction with the environment, by means of a "resonance" and "dissonance" test between percepts and images/concepts, which, together with a mechanism of consolidation of adaptively efficient patterns, enacts the detection and selection of those mental objects which are more adequate to reality (p. 174). This account should be able to neurologically explain the storage of privileged patterns of behaviour, as well as the association between "similar" mental objects, leading to a "grammar" of mental objects (p. 175). Changeux triumphantly concludes: "Darwin allows to reconcile Fodor and Epicurus" (p. 176).

Intentional "inexistence" of mental objects is also explained within this framework: imagination is reduced to the neural "simulation of a future behaviour in front of new situations", which can eventually be "selected" by the global "system of surveillance". In the end *intentionality*, as the capacity to represent, is reduced to a property of the neural system in its (actual or simulated) interaction with the environment, while *consciousness* is *identified* with the system of regulation of neural activity (p. 211). The indistinction of consciousness and

[27] This is the object of a number of technical papers, such as J.-P. Changeux / P. Courrège / A. Danchin, *A Theory of the Epigenesis of Neural Networks by Selective Stabilization of Synapses*, in *Proceedings of the National Academy of Sciences of the United States of America*, 70 (1973), n. 10, pp. 2974–2978. The basic ideas are outlined in Changeux, *L'homme* cit., pp. 276-280. In this book Changeux also declares to be indebted for these ideas to the work of other scientists, including Donald Hebb, René Thom and Gerald Edelman (Id., *L'homme* cit., p. 173).

neural network offers an exemplary view of Changeux' materialistic work hypothesis.

In order to justify this whole research program, Changeux presents materialism as a heuristic necessity. Even though we do not "explain everything", there is no way but "to throw a ladder on the walls of the Bastille of the mental. The 'spiritualist' alternative has been repeatedly set out. Our choice, contrary to the latter, is open to experience, it stimulates a research" (p. 210). Nonetheless the objectives of this research are grounded on clear metaphysical tenets, and therefore are said to be confirmable by future empirical observations. With the perfection of the "ideography" grounded on neuroimaging techniques "it is not utopic to prospect that the image of a mental object will appear one day on a screen" (p. 209). In the last pages of his book, indeed, Changeux declares his support for a materialistic theory of mind, where the recognition of different "aspects" of the same process does not justify the reciprocal isolation of psychological and neurological descriptions (pp. 334-335).

In his later dialogue with Paul Ricoeur on the foundations of brain science, Changeux comes back to the philosophical background of his views and defends a materialistic interpretation of Spinoza, which is coherent with the one advanced in the French Enlightenment and has been empirically corroborated by the neurobiological conception of the brain as "projective system"[28]. This makes clear that Changeux is no eliminationist – and this marks his distance from Dennett and the Churchlands – albeit he does not hold necessary to develop a scientific account of phenomenal consciousness – and this marks his distance from Searle and Chalmers. His theory, by resorting to functionalist explanations, appears close to naturalistic accounts of the kind supported by Millikan and to Putnam's "liberalized functionalism", though his insistence on the possibility of an electro-molecular description of basic representations spells out his sympathy for a full-fledged materialistic theory. A quite different interpretation of very similar scientific hypotheses is provided by Gerald Edelman.

4.2. Edelman: emergentism and Newtonianism

In a famous series of books and a number of technical papers, Edelman sets out to develop an explanation of all major mental properties in a biological framework and thereby "complete Darwin's program"[29]. His account mainly regards

[28] J.-P. Changeux / P. Ricoeur, *Ce qui nous fait penser. La nature et la régle*, Odile Jacob, Paris 1998, pp. 16-17, 56.

[29] See G. Edelman, *Wider than the sky. The Phenomenal Gift of Consciousness*, Yale University Press,

the basic "primary consciousness", defined as the "ability to represent a scene", which man shares with speechless animals and includes the property of intentionality (pp. 119-120). Similarly (and parallel) to Changeux, Edelman distinguishes a general "theory of the selection of neuronal groups" (TSNG), which is meant to account for basic cognitive functions, without involving consciousness, from a hypothesis on the neural correlates of conscious representations (including intentional states), which Edelman, sharing the American philosophical jargon, calls *qualia*.

Edelman's TSNG is grounded on three empirical principles (pp. 39-41)[30]: (a) *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; (b) *Experiential selection*: a continuous process of synaptic selection, occurring within the diverse repertoires of neuronal groups, which is regulated by "value" signals that arise from the activity of the ascending systems of the brain and is continually modified by successful output; (c) *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 the 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 connection of "value-category memory" and "perceptual categorization", enacted by "corresponding" functional clusters of neurons, first produced consciousness as a "remembered present" (p. 55). 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 dynamical nucleus is defined by means of a measure of "neural complexity", grounded on the statistical theory of information. A quantitative threshold helps to draw a distinction between conscious and non-conscious processes. Thereby the hypothesis is connected with mathematical models and design experimental tests[31].

New Haven 2004, pp. 1-3. I consider here only the most recent expositions of the theory (starting from G.M. Edelman / G. Tononi, *A Universe of Consciousness. How Matter Becomes Imagination*, Basic Books, New York 2000), which probably take into account some philosophical criticism of previous, more technical expositions, including a critical review by Searle (see below § 5).

[30] In this section I follow with slight modifications the longer account included in *Mechanism and Phenomenon of Consciousness. Remarks on Models and Ontology in Dennett and Edelman*, in L. Magnani (ed.), *Model-Based Reasoning in Mathematics and Natural Science*, Springer, Berlin 2013 (in press).

[31] This theory is presented with more technical details in Edelman / Tononi, *A Universe of Consciousness*, pp. 125-138.

Edelman's terminology is not very strict about the relation between brain and consciousness: conscious processes "emerge" from neural processes, the latter "entail" or "give rise to" conscious properties by means of the "phenomenal transformation" that results in *qualia*; *qualia* "reflect" neural differentiations. Edelman basically holds that consciousness is a process, whose structural properties can be traced back to structural properties of its material substratum. These words immediately evoke emergentism and epiphenomenalism, but Edelman corrects this view in several ways, presenting an ingenious way to settle the issue of sub-personal and personal level of explanation by means of a classic scientific methodology. First, he stresses that qualia "exist" as evolutionary products and the scientist has to account for their function. He therefore does not consider the possibility of a separation of functional from phenomenal level (as in the mental experiments about "zombies"). All conscious properties must have played a role for the environmental adaptation of the organism to the environment, and therefore they all have an intentional aspect, at least from a phylogenetic point of view.

This identification of intentionality with the activity of the brain and a means of monitoring body-environment interaction is close to many naturalistic philosophers (including Dennett). But since Edelman believes in the "reality" of qualia, he presents several arguments against strictly reductionist programs, which make his theory very close to Searle's "biological naturalism". The relation of *foundation* and *dissimilarity* between the neural (C) and the phenomenal (C') levels of consciousness is illustrated by the same analogy of the explosive and the explosion employed by Searle[32]. Edelman also insists on the limits of mathematical models of phenomenal consciousness: the model of the dynamical nucleus is only a statistical representation of highly complex and constantly changing neural networks, whose modification, depending on the interplay with the environment (which in humans includes itself complex signs, whose meanings depend on social interactions), cannot ever be described in a deterministic manner. In this perspective *qualia* are themselves a temporary model of C, produced by nature, which aid organisms to master this incomputable complexity by introducing phenomenal dimensions and connecting categories to complex data. Indeed some of the conscious properties are "ambiguous (and occasionally irreducible)" and this feature explains the ability of men to creatively elaborate scientific hypotheses about the world[33].

[32] "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" (EDELMAN, *Wider than* cit., p. 63).

[33] G. EDELMAN, *Second Nature. Brain Science and Human Knowledge*, Yale University Press, New Haven 2006, p. 146: "Although it is true that a scientific description of the world hews more closely to

Again, this reminds of Searle's view. Indeed Searle appreciates 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"[34].

In *Second Nature* Edelman replies to the charge of not having explained the "actual feeling of a quale" – without reforming to Searle – by insisting on the limits of standard scientific methodology. 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. Edelman considers consciousness as a matter of experimental evidence and lawlike mathematical description, whose description may be improved, but which is not therefore a mystery. Here is Edelman's "Newtonian" reply:

> "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" (p. 146).

This reflection deserves to be meditated, as it introduces a non-speculative alternative to dualism and monism in the spirit of modern natural science (see § 6), while at the same time conceding the possibility of a scientific treatment of phenomenology, which is not merely functionalistic. It may also express the need to counter the proliferation of speculative hypotheses in contemporary neuroscience, which is evident in Damasio's Spinozistic theory.

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 dynamics leading to such properties *are* scientifically describable, even if the properties themselves cannot be fully reduced".

[34] SEARLE, *The Mystery of Consciousness* cit., pp. 48, 50. Searle's essay is a revised version of a review of Edelman's previous presentations of his theory in the books *Remembered Present* (1989) and *Bright Air, Brilliant Fire. On The Matter of the Mind* (1992), which has appeared in *The New York Book Review*, November 16, 1995.

4.3. Damasio and Spinozism

Since his first popular book *Descartes' Error* (1994) Damasio has always been very sensitive about the philosophical (and metaphysical) foundations of neuroscience, and his theoretical works include plenty of references to both classic and contemporary philosophers. Similar to Edelman, Damasio's account of intentionality is included in his treatment of the "problem of consciousness", which regards the explanation of two different processes: the formation of "images of an object" (representations) and the formation of the sense of ownership, or "the Self". By far the largest part of Damasio's arguments regard the second problem, but the two are deeply connected since consciousness and object-representation emerge as correlative terms of a single intentional process: self-consciousness is defined, from a phenomenological point of view, as "the feeling of what happens when you are being modified by the acts of apprehending something"[35]. From the biological point of view, this feeling corresponds to a mapping mechanism of the brain, which has to be considered as the response of the organism to the interaction with external objects: "core consciousness", as the feeling of being here and now, consists in "neural patterns that map each of the two players [organism and object] and how they interact" (pp. 133, 169)[36].

Intentional representation of external objects is thereby strictly interconnected to consciousness (both in humans and speechless animals). Since the interaction with objects modifies the body of the organism, and the account of this interaction is the relevant information from the point of view of orientation and adaptation to the environment, the object-mapping function is inseparable from the mapping of the body state. This "multidimensional brain representation of the body proper", in turn, is considered as a constantly changing "protoself" (p. 146), that forms the necessary background for the formation of consciousness. The third mapping function, which regards the interaction between objects and body, produces the "image of a relationship" (p. 178) between objects and the proto-self, leading to proper "core consciousness", which is considered as an "inner sense". This view is "curiously [...] no longer mainstream", but is traced back to a number of eminent thinkers of the past: "Locke, Kant, Brentano, Freud,

[35] A. Damasio, *The Feeling of What Happens. Body and Emotion in the Making of Consciousness*, Harvest Book, Orlando 1999, p. 10.

[36] Even though Damasio inserts his hypothesis in a evolutionary framework, and considers consciousness as a tool for survival, he does not beat Changeux' and Edelman's track of developing an evolutionary and dynamic account of the brain structure, but rather looks for a localization of neural correlates of the basic mapping functions (presenting different hypotheses). On the other hand, Damasio positively mentions the hypothesis of detecting a "critical" level of brain mapping as a threshold of consciousness (see e.g. A. Damasio, *Looking for Spinoza*, Vintage Books, London 2004, p. 208).

and William James" (p. 126). As a matter of fact, Damasio joins forces with a vast assembly of modern thinkers who have related the sense of the self to the body, and eventually suggests that this rooting of the self can be considered as a neurologically updated form of Spinoza's prophetic definition of mind as "the idea of the body"[37].

In this framework there is place for the imaginary representation of non-present or "inexistent" objects: it is a "simulation" of physical states happening in a "looped circuit" that excludes the mapping of the actual body state, whose function is to anticipate possible motor reactions to environmental stimuli. In his most recent book Damasio considers this idea, first presented in *Damasio's Error*, as empirically confirmed by the successive discovery of the "mirror neurons" by Giacomo Rizzolatti[38].

One problematic feature of this intriguing account is Damasio's talk of "images", "maps" and "representations" as *conditions* of consciousness, which do not presuppose, but rather *produce* the sense of a representing self. Indeed, Damasio's idea is that neural maps regard *causal* interactions between body and environment, and thereby give rise to a *feeling* and sense of ownership. In *Self Comes to Mind* Damasio counters the objection about how this feeling may arise from purely material mappings with a bold move, that immediately qualifies his view as a modern version of Spinoza's "double-aspect" monism: he *identifies* mental images and maps (p. 65). This coherent monistic move has the advantage of dissolving in a stroke the whole debated issue of personal vs subpersonal level of explanation of cognitive neuroscience and to ground mental representations in the biological process of regulation of the homeostasis. Reminiscent of Spinoza's admonition (in the *Ethics*) that the causal power of the body may well reflect that of the mind – if only one could explain its functions on the ground of the empirical knowledge of its structure [39] – Damasio is therefore free to use an intentional language: the brain "does" the consciousness and has an "intrinsic 'aboutness'" in Brentano's sense, or "intentional attitude regarding the body" (p.

[37] This connection is developed in Damasio, *Looking for* cit., pp. 209-217. A reference is also made to Whitehead (p. 308 footnote). A longer list of related authors includes Kant, Nietzsche, Freud, Merleau-Ponty, Johnson and Lakoff, Edelman, Humphrey, Rosenfield (*The Feeling of What* cit., p. 347 footnote).

[38] A. Damasio, *Self Comes to Mind. Constructing the Conscious Brain*, Pantheon, New York 2010, p. 94.

[39] See B. Spinoza, *Ethica more geometrico demonstrata*, Book III, Prop. II, Note: "No one hitherto has gained such an accurate knowledge of the bodily mechanism [*fabrica*], that he can explain all its functions; nor need I call attention to the fact that many actions are observed in the lower animals, which far transcend human sagacity, and that somnambulists do many things in their sleep, which they would not venture to do when awake: these instances are enough to show, that the body can by the sole laws of its nature do many things which the mind wonders at" (transl. by R.H.M. Elwes, available at http://www.ethicadb.org).

90); it "informs itself", human cultures (as higher products of a basic homeostatic function) are the product of the effort of the "brains". The price paid for this move is very high, since on this ground the whole theory suffers from a blurring of the phenomenological distinction between conscious and not conscious images, mind and brain: mental "images" have an experienced side, but there are also unconscious images (notice how this problem reminds of the post-Cartesian account of obscure ideas).

In *Self Comes to Mind* Damasio recognizes that his theory of consciousness must be integrated in some way (p. 242) and concentrates on a more complex account of the neural network, claiming that it can be useful as a "work hypothesis" in the search for neural correlates of cognitive and affective functions in patients with localized lesions and pathologies (pp. 242-243). But putting this standard kind of inquiry in a monistic metaphysical background still (and unsurprisingly) produces problems at the moment of its application. One of Damasio's more insisted points is that the brain stem nuclei "in all likelihood [...] is the place where the process of making mind begins, in the form of primordial feelings" (p. 243). Pursuing the beginning of mind in restricted brain areas Damasio is led to recognize a "protophenomenal" aspect in single neurons and argues that the emergence of *qualia* from the circular interplay of nervous system and the body suggests a "functional fusion of bodily and perceptive states". Neurons imitate life so thoroughly that they "become one with it" (p. 257). This paves the way to a vitalistic hypothesis that reminds of XVIIIth century materialism: cells appear to possess a sort of "protocognition" (p. 252), which grounds the brain's "aboutness", and they may also possess a kind of "protofeeling" grounded in their "irritability": "this line of inquiry deserves to be pursued" (p. 258)[40].

Without going further into Damasio's conjectures we can isolate a further problem of this metaphysical framework. It must be remembered that Dennett, in his own attempt at dissolving dualism, has decomposed intentionality into non-intentional sub-personal systems:

> "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".

[40] Compare how Maupertuis argued that "the smallest parts of matter" may possess "some principle of intelligence, something similar to what we call desire, aversion, memory". P.L. Maupertuis, *Système de la nature*, in Id., *Oeuvres*, Jean-Marie Bruyset, Lyon 1768, vol. I: § XIV, p. 147 and § XVIII, p. 149.

In a recent reply to critics, Dennett acknowledges that in his account it is not clear at which stage we miss full intentionality, and therefore we can attribute by analogy quasi-intentional states to subsystems of organisms. As Dennett puts it, "we don't attribute fully fledged belief (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". 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"[41].

The problem, now, is that we cannot *empirically* tell the difference between Dennett's and Damasio's "quasi-intentionality", though the two theories presuppose starkly contrasting ontologies: the former denies intentionality as a real ontological property, and reduces it to a theoretical "stance", while the latter admits of consciousness as a fundamental property of the organism and even attributes a minimal part of it to single cells. This is just one example of how the methodological and metaphysical foundation of neuroscience gives rise to interpretative problems when it comes to the explanation of intentionality and consciousness. The materialism, Newtonian phenomenalism and Spinozism, adopted respectively by Changeux, Edelman and Damasio, have presented cases of self-interpretation in neurobiology. We will review now some general solutions offered by philosophers to the problem of orientating and interpreting neuroscientific research.

5. Methodological critiques to the neurobiological approach

A result of the interdisciplinary exchange between philosophy of mind and neuroscience has been the sharing of some pieces of theoretical background, and we have seen how intentionality and *qualia* have become in the last decades accepted terms of scientific inquiry. The emergence of theoretical proposals in the field of neurobiology has been among the causes of a shift from connexionism, grounded on the work on AI and abstract neural networks, to a new trend of natural philosophy. But this kind of exchange is not considered universally to be fruitful among philosophers. The evaluation of this shift can be very different, depending on the philosophical perspective being used.

[41] D. Dennett, *Philosophy as Naïve Anthropology*, in M. Bennett / D. Dennett / P. Hacker / J. Searle, *Neuroscience and Philosophy*, Columbia University Press, New York 2007, pp. 87-88. Dennett draws the last quotation from his earlier paper *Dennett, Daniel C.*, in S. Gutterplan (ed.), *A Companion to the Philosophy of Mind*, Blakwell, Oxford 1994, p. 240.

According to Dennett this contamination of neural science has been for the worst. In the *Preface* to the second edition of *Content and Consciousness* (1985) he considers "very gratifying" the "recent upsurge of enthusiasm among neuroscientists for theories of learning as intracerebral evolution", quoting the work of "Edelman at Rockefeller, Changeux in Paris, the 'New Connectionists' in A.I. and others"[42]. On the other hand he considers qualia as a "poisonous gift" of philosophy of mind to neuroscience and dismisses as "reactionary" the different speculations about new physical of psychophysical theories, which involve a deep modification of standard scientific knowledge in order to account for a first-person ontology[43]. As the quoted list makes clear, Dennett appreciates the work of Edelman for his theory of selection of neuronal groups, while passing over the theory of the dynamical nucleus.

Indeed, in an exchange with Searle, he insists that his instrumentalist method of "heterophenomenology" is "tacitly endorsed" by Edelman and everybody working in the neurological study of intentionality and consciousness, which is realized without any use of "ontological subjectivity". He considers Searle's dissatisfaction with Edelman's theory as a confirmation of Searle's mistaken requirement of a "first-person" theory, something that "no scientific theory" can adopt. According to Dennett, due to this conceptual rejection of standard scientific theory, Searle formulates "no research program" and hides a conservative metaphysical standpoint under the disguise of naturalism. Searle rightly replies that Edelman does not share Dennett's instrumentalist view of consciousness[44]. However, Dennett is also right to emphasize a misunderstanding in Searle's criticism of Edelman and point out an exaggeration in Searle's dismissive attitude towards contemporary neurobiological theories. If Edelman's theory does not explain "exactly how" consciousness emerges, then one may wonder what else must be done in order to satisfy Searle's search for the "causal power" of the brain. Confronted with Edelman's "Newtonian" foundation – which may easily be an implicit reply to Searle – this resistance toward accepting a theory which is close to his ideal of biological naturalism appears to confirm Dennett's malicious critique to Searle as a conservative "mysterian". More generally, it is the sign of a theoretical uncertainty which affects the search for naturalistic ac-

[42] Dennett, *Content and Consciousness* cit., p. xi.

[43] Dennett, *Philosophy as Naïve Anthropology* cit., p. 75. D. Dennett, *Sweet Dreams. Philosophical Obstacles to a Science of Consciousness*, MIT Press, Cambridge Mass. 2005, pp. 8-13.

[44] Searle, *The Mystery of Consciousness* cit., p. 118 (Dennett's reply) and p. 125. Actually Searle considers Dennett's theory to "deny" consciousness in an eliminationist perspective. The point of dissent is the exact meaning of Edelman's work in AI, and both authors refer to Edelman having confirmed their interpretation in "personal communication" – a very significant example of the philosophers' dependence on scientific accounts, whose philosophical meaning may be in itself ultimately ambiguous.

counts of mental phenomena after years of interdisciplinary interplay: scientific accounts may be well founded, but there is no shared view about how they should be empirically confirmed or integrated.

This uncertainty is reflected by the high degree of metaphysical speculation and talk of "plausibility" characterizing both neuroscience and philosophy of mind in the last 20 years. While neuroscientists advanced so far in the field of philosophy, evoking metaphysical systems of the past (think of Damasio), philosophers such as Chalmers can step back from the interpretation of empirical theories to a classification of different forms of metaphysics (see above footnote 22). Philosophers no longer search for a *confirmation* of theories of representation and consciousness in theories about the neural correlates of cognitive functions, even though they recognize the importance of neuroscientific progress; as philosophically engaged neuroscience no longer puts strong constraints on the interpretation of data, and several leading scientists refuse reductionism, more naturalistic philosophers seem convinced that philosophical problems can be solved *with the help of* neuroscience, but not *by* neuroscience.

A different way to exit this unstable condition, and avoid the return of uncritical metaphysical speculation produced by the problems of intentionality and consciousness, has been advanced by so-called "neurophenomenology". This research program derives from the work of Francisco Varela and has received a growing attention in contemporary cognitive science. One of its claims is the necessity of a new methodological foundation of neuroscience. Varela basically highlights the heuristic primacy of phenomenological intuitions and descriptions – first of all regarding intentionality itself as the structure of any mental act – with respect to empirical research on brain tissues and behaviour. In order to know that we are investigating the neural correlates or behaviours that correspond to a given experience, we have to be acquainted with this experience. But neurophenomenology adds a substantial point to this claim, which as such is already widely recognized by contemporary authors, and this addition explains the apparent oxymoron in the title of this research program: we can and must elaborate phenomenological descriptions of experience, in the sense that Husserl expounds on the basis of the phenomenological reduction, because only thereby we can articulate the data for our empirical, naturalistic investigations, without entering the problems of introspectionist psychology. On this crucial point Varela contests Dennett's confusion of the two approaches[45]. Given the phenomenolog-

[45] F. Varela, *Neurophenomenology: A Methodological Remedy to the "Hard Problem"*, in *Journal of Consciousness Studies*, 3 (1996), pp. 334, 338-339 (subsequent references to this paper are given in brackets). Varela's proposal has been originally advanced in the 1970s and later elaborated by a group of adherents to his view. The research program can be articulated in different ways: for instance, Jean Petitot

ical description, the different kinds of intentional states will be successively connected with the neural networks that constitute their material substratum, in order to verify, falsify, or better understand and refine the given articulations.

Neurophenomenology therefore substitutes the raw sense data of Anglo-American empiricist tradition with the intuitive examination of *Erlebnisse* developed along the tradition of European phenomenology. This move has a major consequence for the ontological understanding of intentionality and mental states in general. The idea of regional ontologies, first developed by Husserl on the ground of noematic distinctions and valid in the horizon of pure (non-naturalistically interpreted) phenomena, is considered by Varela as a necessary and complementary addition to the naturalistic world-view, which preserves the irreducibility of consciousness argued by Searle, Chalmers and others. Indeed neurophenomenology recovers Husserl's transcendentalist intuition thesis that the intentional structure of knowledge precedes the distinction between "subjective" and "objective" phenomena (pp. 339-340). Neuroscientific analysis is therefore grounded on a principle of "phenomenological circulation" (pp. 341-345), grounded on the work hypothesis that "phenomenological accounts of the structure of experience and their counterparts in cognitive science relate to each other through reciprocal constraints" (p. 343). The methodological approach of phenomenology allows to acknowledge the importance of both mental experience and brain physiology by considering first-person accounts as a substantial part in the process of confirming a neurobiological hypothesis (p. 344), without entering into any monistic or dualistic ontological claim. This allows to reconcile approaches advanced by different authors (including Searle and Chalmers) whose "naturalistic stance" appears as an unnecessary source of metaphysical problems.

The ideas of neurophenomenology are rooted in the tradition of French phenomenology, in particular in the pioneering work of Merleau-Ponty and Paul Ricoeur on the connection of phenomenology with biological and psychological sciences. Ricoeur's confrontation with neuroscience, which has been conducted all along his philosophical career, has led the philosopher to a direct dialogue with Jean-Paul Changeux on the methods and ontology of the discipline. Here Ricoeur refutes the post-Cartesian opposition between substance-dualism and substance-monism, arguing for a "semantic dualism" of the discourse on the

argues that, in the light of contemporary theories of mathematical physics which were not available to Husserl, one can mathematically describe the emergence of phenomenal forms and thereby return back from the phenomenological level of reduction to a naturalistic account of perception (J. Petitot, *La réorientation naturaliste de la phénoménologie*, in *Archives de Philosophie*, 58/4 [1995], pp. 631-658). For an introduction to the neurophenomenological perspective, which insists on the continuity with the phenomenological tradition, see S. Gallagher / D. Zahavi, *The Phenomenological Mind*, Routledge, Abingdon (Oxon) 2008.

brain and the self, in order to avoid a "semantic amalgam" leading to mistaken expressions such as: "the brain thinks"[46]. In contrast with Changeux' positive remarks on the naturalization of intentionality (p. 37), Ricoeur argues that phenomenological description of the "body proper" or "lived body" (*corps vécu*) can help to think the unity of man in a "mixed discourse" (p. 47). Ricoeur considers all the empirical discoveries of neuroscience and cognitive science about function and organization as relative to the "substratum" of the nervous system, as a material basis of experience (pp. 60-61); but he denies Changeux' thesis that the brain can be in itself "projective" (p. 59), develop consciousness (p. 66), contain images (p. 68), holding that the understanding of experience (both theoretical and practical) is grounded on the fundamental intentional representation of ownership (pp. 25-26) and capacity (p. 42), as a condition of possibility of the body proper. Starting with these notions, which are derived by Husserl's *Ideen II* and Merleau-Ponty's seminal interpretation of this book in the *Phénoménologie de la perception*, Ricoeur advances a distinction between neurological knowledge, which can provide more "explanations", and phenomenological description, which ultimately grounds biological, objective explanations on the background of phenomenological "understanding" of life.

The irreducibility of neurological and phenomenological discourses, advocated by Ricoeur, may appear to suggest no solution to the issues of interdisciplinary work on intentionality, and indeed the dialogue with Changeux does not lead to an agreement (the latter insists on his ontological monism and the unlimited explanatory power of neurobiological approach even in aesthetical and ethical issues). Yet a more sympathetic attitude towards a similar claim is expressed by Edelman in his book *Second Nature*, by addressing the classical issue of natural and cultural sciences. After having denied the possibility to fully explain the phenomenal content of experience, Edelman here considers semantic ability as grounded in the basic process of the production of "ambiguous" properties, which are the ground of metaphorical thinking and give rise to scientific hypotheses and more creative representations. In spite of the fact that "the brain structures and dynamics leading to such properties *are* scientifically describable", Edelman argues that "the properties themselves cannot be fully reduced"[47]. Since Edelman denies that this irreducibility derives from a present limitation of neurobiological description, and argues for the theoretical possibility to describe any mental event in neurological terms[48], with this irre-

[46] Changeux / Ricoeur, *Ce qui nous fait penser* cit., pp. 24-26.

[47] Edelman, *Second Nature* cit., p. 146.

[48] Edelman, *Second Nature* cit, 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 in-

ducibility claim he apparently refers to a different *modality* of description. Thereby semantic and intuitive capacities, that is the two main directions of intentional acts, are presented as the highest product of the brain complexity and a way is shown for the connection of neurobiological knowledge to the description of the "second nature", constituted by the world of human life and culture.

A different argument supporting the need of a methodological re-foundation of neurosciences has been advanced by Bennett and Hacker. These authors have argued that practically all the researchers in the field are misled by a fundamental fallacy, which the authors define following the inspiration of Wittgenstein's argument about intentional ascription and a previous logical analysis of this idea by Anthony Kenny. This "mereological fallacy" consists in the attribution to the brain, or to brain parts, of intentional states, which only makes sense to attribute to the "human being as a whole"[49]. On the light of their criticism neuroscience can indeed empirically investigate the "neural conditions" and "concomitants" for the "possession and exercise" of psychological (including intentional) powers, but this does not affect the "conceptual truth that these powers and their exercise in perception, thought, feeling are attributes of human beings, not of their parts – in particular, not of their brains" (p. 3) and therefore does not allow any talk about the thinking brain or the intentionality of subpersonal systems.

In the course of an animated debate with the authors, Dennett has remarked that this approach merely defends a sort of common sense psychology, by invoking (and not spelling out) a set of standard rules of use, and thereby contradicts with a "obscurantist" attitude the perspectives that have been fruitfully adopted in whole fields of research[50]. Indeed this approach leads to the elimination of entire philosophical objectives in cognitive sciences, which are considered misconceived, while leaving open the fruitful study of biochemical interactions and neural networks in the brain. A Wittgensteinian critique of private feelings and defence of ordinary language substantiate Bennett's and Hacker's critique of neuroscientific talk about intentionality. To Dennett's critique they reply their different *interpretation* of scientific results is ultimately ground-

dicator 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".

[49] M.R. Bennett / P.M.S Hacker, *Philosophical Foundations of Neurosciences*, Blackwell, Malden MA 2003, pp. 68-72. Subsequent references to this book are given in brackets. For the quoted sources of this argument see: L. Wittgenstein, *Philosophical Investigations*, Blackwell, Oxford 1953, § 281; A. Kenny, *The Homunculus Fallacy*, in M. Grene (ed.), *Interpretations of Life and Mind*, Routledge, London 1971, pp. 65-74.

[50] Dennett, *Philosophy as Naïve* cit., pp. 81-86, 90-95.

ed on a precise philosophical and metaphysical tradition: they intend to contrast the mistaken "Cartesianism" of all localization and reduction attempts (pp. 29, 43-46) with an Aristotelian theory of "mental faculties", which avoids the mereological fallacy affecting modern neuroscience and philosophy of mind (p. 15). Thus, their critique constitutes a very relevant example of the metaphysical contrapositions that take place on the backstage of discussions on contemporary neuroscience.

6. Concluding remarks: the re-emergence of metaphysical frameworks and the paradox of "naturalized intentionality"

Among the few points of agreement among philosophers and neuroscientists currently engaged in the naturalization of intentionality – beyond the positive evaluation of empirical research and modelling – is the anti-Cartesian stance. The opposition to Cartesian substance dualism is a common element of all major books in the field, from Dennett to Searle, from Paul and Patricia Churchland to Chalmers, from Edelman to Damasio. Without going into the details of (but presupposing) past criticism of rationalistic metaphysics, and presupposing a more or less articulated naturalistic world view, almost all scholars agree that dualism produces an impediment to research. But Descartes' uninterrupted presence in contemporary discussions suggests that his legacy amounts to more than the misconceived error of hypostatizing thought. Indeed, as it has been occasionally recognized, the very problem of naturalizing intentionality, and the ways through which this problem is supposed to be solved, are conceived on the basis of theoretical assumptions that can be traced back to Descartes' philosophy.

First, the very idea of identifying all kinds of mental states with conscious "thought" derives from Descartes' metaphysics[51]. This is very clear in Searle, as he writes – echoing Descartes' argument in the *Meditations* – that "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*"[52]. Here is also the ultimate source of the conception of intentionality as a conscious process, which forms a background notion for the very formulation of the problem of naturalizing intentionality (and consciousness).

[51] Descartes considers "thought" (*cogitatio*) as the essence of mind and defines thought, in turn, as "everything which takes place in us so that we are conscious of it", including "to think, to will, to imagine and also to feel". R. Descartes, *Principia philosophiae*, I, 9, in C. Adam / P. Tannery (eds.), *Oeuvres de Descartes*, Vrin, Paris (1964-74)^2, VIII, p. 7.

[52] Searle, *The Mystery of Consciousness* cit., p. 112.

Secondly, the idea of the body as a machine and of brain "mechanisms", as composing the reality which grounds mental faculties and causes behaviour, is in itself a Cartesian claim, which was originally limited to the account of involuntary processes, and has inspired research on the physiology of mental states from the XVIIth to the XXth century and beyond. Most neuroscientists recognize this historical role of Descartes[53], while disagreeing on whether to pursue or not a consequent, physiological account of intentionality without any reference to consciousness and feelings.

The persisting influence of these Cartesian tenets is the reason why dualism is still considered a challenge for the dominant naturalistic perspective of contemporary philosophy. The search for alternative naturalistic frameworks, including the postulate of an inner reform of scientific theory, is a way to contrast this challenge. Nonetheless, as our case studies have shown, many philosophers and scientists think that no empirical results seem to be in the position to solve the problem, without a concomitant reform of Descartes' concept of the mind. An alternative solution to the same problem is to draw a distinction between different and irreducible levels of description. This kind of solution, which is cautiously suggested in Putnam's liberalized realism, is adopted by phenomenological approaches, which thereby allow the distinction between the irreducibility of a non-naturalistic level of description from metaphysical issues. According to Ricoeur, indeed, Descartes' late reflections on ownership of the body in the sixth *Meditation* and his later reflections on a "third substance" (pp. 39, 51-53) can be considered as the modern historical root of a phenomenological understanding of life, which was rendered finally possible after Kant's criticism of dogmatic metaphysics of substance (p. 25).

This situation also explains the interest in Spinozism, as an exemplary monistic solution to the metaphysical problems of the Cartesian Age. Nonetheless the conflict between naturalism and phenomenology also conditions the reappraisal of Spinozism. Changeux reads the *Ethics* with the glasses of XVIIIth century radical materialism, considering the latter as the most coherent prosecution of Descartes' science of man; in a similar way Damasio invokes Spinoza's monism as a background to reconsider the hypothesis of attributing "proto-mental" properties to cells. Ricoeur, on the contrary, considers Spinoza's monism as a consequent ontological settlement of Descartes' issue of the living body (pp. 30-31, 34). This conflict of interpretations may alternatively be seen as the sign of irreducible differences and as a means of theoretical cooperation. Its main problem

[53] Changeux, *Ce qui nous fait penser* cit., p. 47-54 insists on the ambiguity of Descartes and suggests that his metaphysics may not contain his serious beliefs. Also see Bennett / Hacker, *Philosophical Foundations* cit., p. 30.

appears to lie in the connection of metaphysics to empirical theories, which actually needs a quite different elaboration of monism compared to Spinoza's original metaphysics; indeed, as it is suggested by Ricoeur's remark on Kant, the whole issue usually lacks a full critical investigation of metaphysical hypotheses, being limited to a denial of dualism.

A similar remark can be made regarding the revival of Aristotelian philosophy, which is also a characteristic aspect of contemporary philosophy of mind. A reappraisal of Aristotle has been suggested as a logical alternative to the shortcomings of Cartesian mechanism and dualism and has led to a number of different accounts of intentionality. By restricting the field to naturalistic approaches, we can think of Putnam's and McDowell's theories of meaning. Putnam's critique of internal representations, for example, is presented as a revival of an originally Aristotelian "direct realism", which is opposed to the "dominant view in Anglo-American philosophy of mind" which "today appears to be what we may call 'Cartesianism-cum-materialism'"[54]. But the main point of contact with the problem of intentionality regards the approach to mind as the bearer of faculties. A first example is provided again by Putnam, whose functionalism is directly traced back to the Aristotelian notion of "form"[55]. This background would allow the embedding of models in cognitive and neural science on the non-dualistic background of a set of elementary functions implemented in the body, and thereby would avoid speculative metaphysical claims about *qualia* and the essence of matter.

This account lacks a more detailed treatment of how aspects of Aristotelian philosophy can be actually connected to contemporary science of nature. First, Aristotle's soul with its faculties is indeed inseparable from the body, but this does not eliminate the problem of connecting the account of logical processes guiding perception and voluntary action with its bodily counterpart, which Aristotle faced with the tools of his physiology. Secondly, some claims of Aristotle's metaphysics, such as teleology or the theory of the intellect, contrast with the presuppositions of contemporary – post-Cartesian and post-Kantian – philosophical naturalism. Moreover the interpretation of Aristotle as a direct realist has been challenged by recent scholarship[56].

A similar problem affects Bennett's and Hacker's Aristotelian foundations of

[54] See e.g. Putnam, *Aristotle's Mind and the Contemporary Mind*, in Id., *Philosophy in an Age* cit., pp. 584-585, 589. Compare Millikan's conclusion that his theory of meaning, "though the starkest possible antithesis of rationalism, will still be close to Aristotelian realism. Properties and kinds will show up only in the actual world. Nominalism will be denied" (Millikan, *Language, Thought and Other* cit., p. 11).

[55] H. Putnam / M.C. Nussbaum, *Changing Aristotle's Mind*, in M.C. Nussbaum / A. Rorty (eds.), *Essays on Aristotle's "De Anima"*, Oxford University Press, Oxford 1992, pp. 27-56.

[56] See V. Caston, *Aristotle and the Problem of Intentionality*, in *Philosophy and Phenomenological Research*, 58 (1998), 2, pp. 249-298.

neuroscience, leading to a significant "watering down" of Aristotelism. The authors formulate the charge of "Cartesian materialism" to the whole discipline, arguing that the attribution of intentional states to brain matter by contemporary scientists and philosophers is logically equivalent (and equally mistaken) to Descartes' attribution of these properties to the immaterial soul[57]. On the contrary, they defend an Aristotelian "paradigm" of the mind as endowed with faculties and associated to the whole animal, and to this view they trace back the origin of their "mereological fallacy" argument[58]. This view can be considered a sort of functionalism, grounded on pragmatic linguistic arguments. Yet Hacker's and Bennett's view is not meant to *ground* and *be specified in* empirical research, but rather to limit the latter's scope, in a way which is more similar to the phenomenological approach than to naturalism[59].

On the whole, the adoption of metaphysical paradigms of the past, whether dualistic or monistic, does not suffice by itself to determine a solution to the problem of naturalizing intentionality. Contemporary programs of naturalizing intentionality apparently move in a loop, or "pendulum swing", which is a characteristic feature of post-positivistic philosophical naturalism. The problem of connecting intentionality to scientific theories produces the resort to metaphysical backgrounds, restoring a typically modern set of alternatives (Aristotelism, Cartesianism, Materialism, Spinozism). But this reference to metaphysical frameworks carries the burden of highly speculative commitments, and natural science, which is still considered as the only ultimate source of limitations to the metaphysical claims, does not provide a sufficient and autonomous criterion to rule out or confirm metaphysical claims about intentionality. The provisional character of science and the role of metaphysical hypotheses in a holistic epistemology, highlighted in the philosophy of science since the times of Popper and Kuhn, suggest that things may go the other way round, and metaphysical frameworks may orientate empirical inquiry towards new theories. As it is exemplarily shown by Edelman's defence, Newton's theory of scientific knowledge can suggest how far a naturalistic theory can be justified to advance in the explanation of intentional phenomena, and thereby set the border with a different kind of description. But Newton's gravity, on the contrary, is invoked by Chalmers in support of his idea of enlarging the lexicon of natural science with new concepts and new laws.

[57] Bennett / Hacker, *Philosophical Foundations* cit., pp. 85-88, 111-112, 261-262.

[58] See M. Bennett / P. Hacker, *The Conceptual Presuppositions of Cognitive Neuroscience*, in M. Bennett et Al., *Neuroscience and Philosophy* cit., pp. 130-133. The authors derive "Aristotle's principle" from the following sentences of the *De anima*, 408b, 12-15: "to say that a soul is angry is as if one were to say that the soul weaves or builds. For it is surely better not to say that the soul pities, learns or thinks, but that a man does these with his soul".

[59] This similarity is evident, e.g., in Bennett / Hacker, *Philosophical Foundations* cit., p. 115.

In front of this hermeneutical conflict biological (and reductionist) accounts of representation still represent the promise to restore natural science in its role of setting a reference term for a "critique of metaphysics". As an alternative to this program, phenomenology – following the Kantian idea of separating fields of experience with their different forms of legality – represents the possibility of keeping together empirical investigation and common experience, without entering an unsolvable conflict of metaphysical views, while on the other hand ruling out some of the objectives of neuropsychology and neuroethics. The ongoing foundation of the notion of "naturalized intentionality" currently faces these historical and theoretical issues in a dynamic interdisciplinary dialogue: the ultimate challenge of this paradoxical inquiry appears to be not the more or less successful blending of intentionality in a given naturalistic framework, but the critical assessment of the meaning and scope of the notions of 'nature' and 'natural science', as they are required to explain the intentional dimension of human experience.

Abstract: To give an account of intentionality in terms of the concepts and methods of natural science has been considered as a crucial step towards a naturalization of mental phenomena in general, and as such it has been pursued by a large number of naturalist philosophers and cognitive scientists. Starting from the late 1960s the problem has been addressed in very different, reductionist (Dennett, Millikan: § 2) and antireductionist ways (e.g. Searle, Chalmers, Putnam: § 3). The development of these philosophical programs has benefited from the contemporary technical and theoretical progresses of neuroscience, and leading scientists such as Changeux, Edelman and Damasio have presented articulated proposals of naturalization of intentionality (§ 4). A common element of philosophical investigations turns out to be the reference to a still undeveloped neuroscientific theory. This reference belongs to the legacy of early XXth century anti-metaphysical "scientific philosophy". In spite of this dominant philosophical approach, neuroscientists present their pioneering researches on intentionality with the help of metaphysical frameworks of the past, including Aristotelianism, materialism, emergentism and Spinozism. The final section of the paper examines this peculiar "paradox" of naturalization programs, by considering some critical exchanges about the neurobiological approach (§ 5) and by reviewing the role of metaphysical paradigms for different approaches to this issue (§ 6).

Key words: Naturalism; Intentionality; Neuroscience; Reductionism; Metaphysics.

Mailing address:
Università degli Studi di Cassino e del Lazio meridionale
Dipartimento di Scienze umane, sociali e della salute
via Mazzaroppi, 6
I - 03043 Cassino (Fr)
paolopatch@yahoo.it