



Giuseppe Vicari

Beyond Conceptual Dualism

*Ontology of Consciousness,
Mental Causation, and Holism in
John R. Searle's Philosophy of Mind*

VALUE INQUIRY BOOK SERIES

vibs

BEYOND CONCEPTUAL DUALISM

Ontology of Consciousness,
Mental Causation, and Holism in
John R. Searle's Philosophy of Mind



VIBS

Volume 196

Robert Ginsberg
Founding Editor

Leonidas Donskis
Executive Editor

Associate Editors

G. John M. Abbarno	Steven V. Hicks
George Allan	Richard T. Hull
Gerhold K. Becker	Michael Krausz
Raymond Angelo Belliotti	Mark Letteri
Kenneth A. Bryson	Vincent L. Luizzi
C. Stephen Byrum	Adrianne McEvoy
Harvey Cormier	Alan Milchman
Robert A. Delfino	Peter A. Redpath
Rem B. Edwards	Alan Rosenberg
Malcolm D. Evans	Arleen L. F. Salles
Daniel B. Gallagher	John R. Shook
Andrew Fitz-Gibbon	Eddy Souffrant
Francesc Forn i Argimon	Tuija Takala
William Gay	Emil Višňovský
Dane R. Gordon	Anne Waters
J. Everet Green	John R. Welch
Heta Aleksandra Gylling	Thomas Woods
Matti Häyry	

a volume in
Cognitive Science
CS

Francesc Forn i Argimon, Editor

BEYOND CONCEPTUAL DUALISM

Ontology of Consciousness,
Mental Causation, and Holism in
John R. Searle's Philosophy of Mind

Giuseppe Vicari

Guest Foreword by
John R. Searle

Editorial Foreword by
Francesc Forn i Argimon



Amsterdam - New York, NY 2008

Cover photo: Matthew G. Hull

Cover Design: Studio Pollmann

The paper on which this book is printed meets the requirements of “ISO 9706:1994, Information and documentation - Paper for documents - Requirements for permanence”.

ISBN: 978-90-420-2466-3

©Editions Rodopi B.V., Amsterdam - New York, NY 2008

Printed in the Netherlands

Cognitive Science
(CS)

Francesc Forn i Argimon
Editor

Other Titles in CS

Oscar Vilarroya. *The Dissolution of Mind: A Fable of How Experience Gives Rise to Cognition*. 2002. VIBS 128

Gábor Forrai and George Kampis, eds. *Intentionality: Past and Future*. 2005. VIBS 173

Oscar Vilarroya and Francesc Forn i Argimon, eds. *Social Brain Matters: Stances on the Neurobiology of Social Cognition*. 2007. VIBS 190

For Nunzia Ferrante Vicari, Mauro Vicari,
Alberto Vicari, and Ludovica Cassibba.

CONTENTS

EDITORIAL FOREWORD: Special Series in Cognitive Science	xi
FRANCESC FORN I ARGIMON	
GUEST FOREWORD	xiii
JOHN R. SEARLE	
PREFACE	xvii
INTRODUCTION	1
ONE	
Problems and Theories: The Mind-Body Problem in Contemporary Philosophical Debate	9
1. Explanatory Gap, Dualism, and Antireductionism	9
2. Problems with Dualism: The Problem of Mental Causation	15
3. Materialism, Functionalism, and the Desubstantialization of Mind	17
4. Problems with Functionalism: Consciousness, Mental Causation, and Incompleteness	20
5. Return to the Roots of the Desubstantialization of Mind: The Mind-Body Problem in John Locke's Philosophy	25
6. Diagnosis of the Reasons of a Dilemma: John Searle and Conceptual Dualism	33
TWO	
Biological Naturalism: A Naturalistic and Non-Reductive Ontology of Consciousness	39
1. Identifying the Target: The Structural Features of Consciousness	39
2. Biological Naturalism	47
3. Emergent Properties	49
4. Mental Causation: Consciousness and the Creative Structure of the Mind-World Relationship	55
5. Reduction and the Practice of Definition	58
6. The Causal Exclusion Dilemma: Terms of the Problem	62
7. Biological Naturalism and Causal Exclusion Dilemma in the Debate between Kim and Searle	68
THREE	
Functionalistic Models of Consciousness: Dennett, Chalmers, and the Desubstantialization of Mind	73
1. Dennett's Eliminative Naturalism: Multiple Drafts Model vs. the Myth of the Cartesian Theater	73

	2. The Reasons of the Phenomenal: Searle vs. Dennett	80
	3. Chalmers's Neodualism	85
	4. The Reasons for the Unity of the Natural World: Searle vs. Chalmers	90
	5. Desubstantialization of Mind, Functionalism, and Conceptual Dualism	95
FOUR	Holism and Mental Causation in the Theory of Intentionality	101
	1. The Logical Structure of Intentional Representations: Psychological Mode, Representational Content, Direction of Fit, and Conditions of Satisfaction	102
	2. The Causal Self-Referentiality of Perception and Action	105
	3. Network and Background: The Holistic and Pragmatic Structure of Mind	109
	4. The Intersubjective Dimension of Individual Consciousness and Intentionality: Problems and First Steps Toward a Theoretical Account	114
	5. Collective Intentionality and Institutional Contexts: Intersubjective Structures of Consciousness	120
	6. Searle's Theory of Intentionality in the View of Contemporary Scientific Research: Action, Perception, and the Background Sense of Other	125
FIVE	John Searle and Contemporary Neuroscience. Holism, Mental Causation, and the Roots of Subjectivity	131
	1. Building Block Approach vs. Unified Field Model: Implications of the Interdependence of Qualitativeness, Subjectivity, and Holistic Unity of Consciousness for Scientific Research	133
	2. Edelman, Tononi, and the Dynamic Core Hypothesis	135
	3. From Consciousness to the Self: Some Possible Post-Humean Eliminative Trends	139
	4. Consciousness as a Space of Self-Referentiality: Damasio's Theory and the Role of the Proto-Self in Organism-Environment Transactions	146
	5. Subjective Ontology and the Construction of a Sense of the Self: The Interdependence of Holism, Ontology of Consciousness, and Mental Causation	152
	CONCLUSION	159
	WORKS CITED	165
	ABOUT THE AUTHOR	175
	INDEX	177

EDITORIAL FOREWORD

VIBS Special Series in Cognitive Science was born in 2002, under the creative spell of its first editor and author, Oscar Vilarroya, who now holds the Social Brain Chair at the Autonomous University of Barcelona, and the enthusiastic support of VIBS founder Robert Ginsberg. Its primary aim was to bridge what we could call the new explanatory gap, closely related to the one that divides the mental and the physical worlds according to many philosophers of mind since René Descartes. What appeared necessary to connect were not just mental and physical phenomena, but separated lines of research about those phenomena, namely conceptual analyses of mind and cognition in the philosophy of mind on one side and the advances in experimental cognitive science on the other.

Six years and four volumes later, Oscar's aim continues to inspire the editorial philosophy of this Special Series in Cognitive Science within VIBS. The only addition, if any, is a greater emphasis on the relevance of fine-grained studies of the texts of earlier philosophers like Descartes and many others, who set the foundations of the main problems in the field, and the space of possible solutions.

The continuity of purpose and the complementarity of the historical based approach are apparent in this monographic volume. It discusses the work of John R. Searle, one of the most authoritative voices in contemporary philosophy, who is particularly concerned with the relations of mind and world. The author, Giuseppe Vicari, displays with depth and clarity of what is at stake in the current philosophical debates on the nature of those relations. He makes a strong case for Searle's contention that we can bridge the original explanatory gap without reducing either side to the other, and roots Searle's insights on that matter in a suggestive interpretation of John Locke's thought.

One of Vicari's goals is to show how Searle's theoretical model explains the results of cognitive science better than its rivals, and that it is also more attuned to what outstanding cognitive scientists say when they try to make sense of their empirical work. Regardless whether the author succeeds in accomplishing this ambitious goal, he sure compels us to recognize Searle's philosophy as an exceptional contribution to the study of mind, and therefore to cognitive science, as a multidisciplinary effort devoted to our understanding.

Francesc Forn i Argimon
Editor, Cognitive Science

GUEST FOREWORD

A common mistaken assumption about contemporary philosophy is that there is a fundamental distinction between philosophy as practiced in English-speaking countries such as Britain, the United States, and Australia, and so-called Continental philosophy. Such a view, I believe, is obviously mistaken, but Giuseppe Vicari provides a clear demonstration of its inadequacy in his excellent book *Beyond Conceptual Dualism*.

The main theme of Vicari's book is the exposition and explanation of my various writings about the mind. I must say that he has understood me and explained my views better than any other author I have read. However, if this book were just an exposition of my views, it would be much less valuable than it is in fact. With remarkable thoroughness, he situates my views in relation to all of the major authors on the subject, who have criticized and discussed my views on the same sets of issues. With patience, thoroughness, and relentless rigor, he analyzes the various points in dispute; with complete clarity and fairness, he adjudicates the force of the arguments on each side. Furthermore, he is not content merely to present the philosophical arguments; he provides a powerful and extensive discussion of current neurobiological research, especially the work of Gerald M. Edelman, Giulio Tononi, and Antonio Damasio. Vicari sees how my work is related to, and supportive of, the best current work in neurobiology.

One of the most profound aspects of the book, indeed perhaps its most profound, is that Vicari fully understands the importance of the idea of "conceptual dualism." He sees that it is not enough simply to attack current mistakes made under the headings of materialism, dualism, mentalism, or reductionism, but also that we must understand the deep and mistaken assumptions that make these disputes inevitable and irresolvable within the framework of the existing conceptual apparatus. He sees, what I think is the most significant idea I have tried to get across, that we will not succeed in overcoming this whole tradition that goes back to René Descartes, and even before, unless we abandon the conceptual assumptions that give rise to it and make it inevitable.

The problem, as Vicari correctly points out, is not just that people mistakenly suppose that if something is irreducibly mental, then it cannot be thereby and in that very respect physical, and if something is irreducibly physical, then it cannot thereby and in that very respect be mental. I argue that this is a mistake, but as Vicari sees, it is part of a whole series of mistaken assumptions about identity, causation, and reduction. If you stick with the traditional categories, you are committed to a variety of inconsistencies. Once you recognize that the mental qua ontologically subjective is ontologically irreducible to the physical, then, in the traditional categories, you are forced to some sort of dualism. Notice the assumptions about reduction, dualism, and mentalism that are being made.

I, on the other hand, want to stay with what I take to be the fundamental facts. If you state the facts, and it turns out that the facts are inconsistent with a series of traditional philosophical assumptions involving the traditional philosophical categories, then I will stay with the facts and reject the tradition. Of the many commentators on my work, Vicari is one of the very few who have appreciated this point.

I want to illustrate what I think is the correct way of dealing with these problems and the inadequacy of the traditional ways by considering in some detail the sort of example that interests me and Vicari. Suppose I now intentionally raise my arm. We can say with a great deal of confidence two things about this event:

- (1) My intention-in-action caused my arm to go up.
- (2) Anything that caused my arm to go up in that way must cause the secretion of acetylcholine at the axon endplates of the motor neurons.

From (1) and (2) it follows that:

- (3) One and the same event has a level of description where it is a conscious form of intentionality, and where it has neurobiological properties sufficient to affect the synaptic cleft.

If we think about how nature works, unquestionably this, or something like it, must be true. The neurobiological details might be refined, but the fundamental causal role of consciousness in effecting neurobiology is not in doubt.

Notice the difficulties that the philosophical tradition has with this simple case. Because consciousness is irreducible to any third person ontology, the tradition has to say that for my conscious event to cause neurobiological events is impossible. To grant such causal relations would be to violate the sacred principle of “the causal closure of the physical.” This leads to the desperate thesis of epiphenomenalism, according to which no mental event in history ever had any physical effect.

One way out of this, from the frying pan into the fire, so to speak, is to say that the conscious event is just a neurobiological event neurobiologically described. No such thing as irreducible first person consciousness exists. This is “from the frying pan into the fire” because in place of the preposterous thesis that consciousness has no causal effects at all, it substitutes the even more preposterous thesis that irreducible consciousness does not exist. Both of these mistakes are direct consequences of the tradition with its vocabulary. The mistakes are not only about “mental” and “physical” but also about “identity,” “causation,” and “reduction.”

To me, the most surprising feature of Vicari’s discussion is the key set of similarities he finds between my views and those of John Locke. I have always admired Locke’s work, but I was unaware that we had similar views

on mind-body relations. I had not previously found this point made in any discussion, but I must say that Vicari presents good arguments to show that my views are quite similar to those of Locke.

I hope that Vicari's excellent and useful book will be one of many to come that will further help overcome the artificial separation between so-called Continental and analytic philosophy.

John R. Searle
University of California, Berkeley

PREFACE

This work stems from eight years of study and research in the fields of history of philosophy and philosophy of mind. It is a refined version of my PhD dissertation in philosophy. The general question underlying my research has been, in its broad outlines: What kind of picture of our being in the world do we get under the assumption of what we know about the world and ourselves from biology, neurosciences, social sciences, psychology, and other scientific enterprises?

A great part of the debate in philosophy of mind, which constitutes part of the above-mentioned issue, appears still deeply linked to a Cartesian conceptual framework. Philosophers require us to choose between materialism (the idea that mind is “nothing but” an objective, physical phenomenon), which appears incapable of “saving the phenomena” of mind as subjectivity and a range of neodualistic positions which, with the claim of a logical or ontological independence of mind and nature, involve the transformation into the mystery of our being a part of the natural world as conscious, mindful animals.

My first reading of John Locke’s *Essay Concerning Human Understanding* and of John R. Searle’s *The Rediscovery of the Mind*, together with Francesca Di Lorenzo Ajello’s lectures on these issues, made me aware that the dilemma between materialism and dualism, far from providing reliable solutions to the mind-body problem, can be considered as the primary source of it.

In quite different times and contexts, Locke and Searle developed a view of mind that overcomes the traditional monism-dualism dilemma. Their crucial point is that we have to see mind not as a Cartesian substance, but as a set of subjective powers, abilities, and processes realized as part of the biological makeup of organisms and responsible for the mediation of the organism-environment transactions.

Subjectivity, in these philosophical perspectives, is a part of the biological life of organisms and is therefore part of nature. Searle and Locke are not reducing or eliminating mind, instead they see mind as a subjective part of nature.

Further inquiries suggested to me that the issue is not a purely philosophical one. The need for a richer view of mind that overcomes the constraints of the materialism-dualism dilemma is also widespread among scientists.

The works of Charles Darwin, Jean Piaget, Antonio Damasio, Gerald M. Edelman, Walter Freeman, Rodolfo Llinàs, and many other scientists suggest that we have to see our powers of knowledge, theoretical and practical reasoning, creativity, and so on—all of them typical philosophical topics—as tools that we use in our fight for survival, developed as extensions of our basic biological makeup. We can read these works as attempts to understand the embodiment of mind and consciousness as a part of the evolutionary history of biological organisms continuously renegotiating their relationships with the environment, where the environment in question is physical and, in the case of human beings, social, institutional, and cultural. The acting organism and the acting brain take the place

of the old Cartesian picture of a substantial and passive subjectivity that merely mirrors the world.

Recent experiments, for example, show that the sophisticated human cognitive abilities and the intersubjective structure of consciousness have their roots in a prelinguistic, preconceptual, and pragmatic understanding of the intentions and actions of other people (Rizzolatti and Sinigaglia, 2006). This primitive ability is embodied in particular areas of prefrontal motor cortex where the same neurons (so-called mirror neurons) firing when a subject performs an intentional action also fire when the subject observes the same action made by other subjects.

The discovery of mirror neurons also enforces the thesis that action is not the result of pure perception and cognition. Instead, these supposedly independent processes are already mediated by those possibilities of action that the organism structures on the ground of environmental features and conceptual-categorical patterns of action acquired in the ontogenetic and phylogenetic history (Piaget, 1972; 1975; Edelman, 1989; 1992)

The idea of a work on Searle's philosophy of mind was based, therefore, on a theoretical need that I found widespread in philosophy and science. I wanted to show, through an analysis of Searle's position, that we can have a coherent naturalistic and non-reductive ontology of mind that can fruitfully dialogue with scientific research and face the traditional philosophical problems.

I also had interpretive and contextual reasons, since as far as I know, at present the literature has no systematic analysis and critical assessment of Searle's philosophy of mind and of its relationships with scientific research.

Many papers, collections of essays, and monographs are available on single topics and on Searle's philosophical thought as a whole, but none of them focus on Searle's philosophy of mind, giving a systematic account of it.

The work closest to this volume is Di Lorenzo Ajello's 2001 book, *Mente, azione e linguaggio nel pensiero di John R. Searle* (*Mind, Action, and Language in John R. Searle's Thought*). It focuses on the relationships among mind, action, and language in Searle's philosophical proposal. In addition, it shows, against Jürgen Habermas and Karl O. Apel's objections, that Searle's theory of intentionality is still a part of the intersubjectivist and linguistic turn of contemporary philosophy. The rediscovery of the mind does not imply by itself a return to René Descartes' dualism and solipsism.

Other books with excellent papers or chapters on Searle's philosophy of mind are *Searle and his Critics* (Lepore and van Gulick, 1991), N. Fotion's *John Searle*, (2000), *Speech Acts, Mind and Social Reality* (Grewengorf and Meggle, 2002), *Views into the Chinese Room* (Preston and Bishop, 2002), and *John Searle* (Smith, 2003). All of these works contribute significantly to the discussion, but none of them provides a general and systematic account, which is not possible in the space of a single paper, an introductory book, or of a book devoted to a single aspect of Searle's philosophy of mind.

Many issues, which I do not deepen in this work, remain open for future research. Among these issues we have the problem of characterizing the relations

between the subjectivity of mind and its intersubjective structure; the problem of the ontology and genesis of the self in relation to the criteria of personal identity; the relations between conscious and unconscious mental states; the relations between Searle's proposals and the new developments in cognitive sciences, such as the "embodied cognition" perspective (Clark, 1997).

These are all valuable issues that influence my research interests. Searle's philosophy of mind can provide a useful framework to shed light on them.

I have discussed the present work and its ancestors in conferences, seminars, and lectures I gave or attended between 2004 and 2007.

Di Lorenzo Ajello was my supervisor during my doctoral studies in Philosophy. I discussed my ideas with her and with my colleagues, Sandro Guli, Claudia Rosciglione, and Giancarlo Zanet at the seminars and meetings of the research group in History of Philosophy of the *Dipartimento di Filosofia, Storia e Critica dei Saperi* (University of Palermo). I owe special thanks to Di Lorenzo Ajello for teaching me the method of research in philosophy, the constant effort in the search for good reasons and arguments for my views, and above all, for teaching me to love the responsibility and the hard work of critical thinking.

Searle read the final version of the manuscript and gave me several suggestions and objections on the entire subject matter of the book, with particular reference to the themes of mental causation, intentionality, and the relation between the self and causal self-referentiality. I owe him special thanks for the many helpful comments, suggestions, and critiques that he gave me discussing my ideas and his work in the occasion of the lectures and conferences that he gave in Berkeley (California, USA, 2005 and 2007), Lublin (Poland, 2005), and Palermo (Italy, 2006 and 2007).

The ideas of Chapter One, section three, on Locke's philosophy of mind, were discussed in lectures and conversations with Yasuhiko Tomida at Locke's Tercentenary Conference (Oxford, St. Anne's College, 2004), Edwin McCann, and Lionel Shapiro (Berkeley, 2005).

The subject matter of Chapters Two, Three, and Four, concerning Searle's philosophy of mind and the debates with Jaegwon Kim, Daniel C. Dennett, and David J. Chalmers were read and discussed at the XVI National Conference of PhD in Philosophy, section on "The Contemporary Debate: Representation, Concept, and Action" (Reggio Emilia, Italy 2005). I had useful suggestions and critiques from all the participants, particularly from Alberto Peruzzi, Paolo Parrini, and Carlo Penco.

Part of Chapter Five was discussed in a seminar I gave at the Social Ontology Group (John R. Searle, Chair) at the *Department of Philosophy* of the University of California, Berkeley. John Searle, Anders Hedman, Jennifer Hudin, Beatrice Kobow, Asya Passinsky, Antti Saaristo, and Trevor Wedman made several comments and critiques of my views on the issue of the self that I took into account in the present work.

Francesc Forn Argimon, Eric van Broekhuizen, Elizabeth D. Boepple, and Asya Passinsky made a crucial contribution to improve the quality of the book

with their constant search for clearness and comprehensibility of the style and the arguments.

My research and work would never have been possible without the moral, economical, and loving support of my family and friends.

Without any claim for completeness, I want to remember Mauro, Alberto, Paolo, Gisella, and Cristian Vicari, Serena Delmonte, Massimiliano and Serena Cassibba, Mariella, Francesca, and Enrico Scaccianoce, Rocco, Sandro Ferrante, Francesco Iellamo, Nicola Nocera, Luisa Adamo, and Anna Maria Barbera. Antonio and Ina Ferrante and Richard T. Hull, did diligent proofreading of the manuscript; for that, I am grateful.

Special thanks to my mother, Nunzia Ferrante Vicari. This book is devoted to her memory.

INTRODUCTION

After the end of the 1980s, the problem of consciousness emerged from the more general philosophical and scientific discussions of the mind-body problem and became the hard problem in philosophy of mind and cognitive science. After the birth of cognitive science and the escalation of philosophical and scientific studies grounded on the paradigm of classical computational functionalism, philosophers and scientists discussed the evolutionary efficacy of consciousness and the related problem of mental causation, on the nature and existence of qualitative mental states, on the real or merely virtual existence of the self, and on emotions and rationality.

A crucial problem overarches these issues: we need an ontology that accounts for the relations between the irreducibly different (but not independent) levels of organization in the natural world within a naturalistic conceptual framework for our analysis of mental phenomena. In the case of mind, we would like to know how mind and consciousness are a part of the natural world without leaving out those features (such as qualitativeness and subjectivity) that make mental phenomena different with respect to other natural phenomena. Ultimately, then, our model should account for the power of some biological organisms to produce and support what John R. Searle calls “conscious fields of qualitative, unified subjectivity” (2004, p. 154).

The debate over this issue, which is a part of the mind-body problem, is in Searle’s view, grounded on the disagreement between materialism (or monism, or physicalism), that argues for the scientific irrelevance or the denial of the existence of the subjective and qualitative character of mind, and a minority of scholars who argued for the irreducibility of consciousness as a subjective and qualitative phenomenon, who consider themselves property dualists or skeptics. The reasons for the difficulty, apparently, are that conscious states are characterized by “new” properties with respect to the rest of the natural world: they have a qualitative dimension, a subjective form of existence (they exist only when a subject experiences them), and they exist only within a network of other mental states, conscious and otherwise (Searle, 1983, p. 19ff.).

Many philosophers made a definite choice between these traditional positions. We have well known examples of materialism with philosophers such as Thomas Hobbes, Willard van Orman Quine, Richard Rorty, Daniel C. Dennett, and many others who think that mental states are “nothing but” physical states. We also have well known examples of philosophers who think that mind is ontologically different from the physical, such as Plato, Gottfried W. Leibniz, Karl R. Popper, and David J. Chalmers.

Others tried to develop a third view to overcome this traditional monism-dualism dilemma and worked out what I here call the “desubstantialization” of mind. According to this view, mind is not a material or immaterial “substance,” but a set of processes dependent on the physical and realized in it, devoted to the

mediation of the organism-environment relations but, *qua* processes, irreducible to the realizing physical system.

This view of mind is rooted in modern, postCartesian philosophy, beginning with John Locke's thesis that mental states are not "faculties" or "substances" but "powers" or operations of an agent. On this account, thinking is not the essence of an immaterial soul, but one among the operations that an agent can perform and the function of which is to enhance the well-being of the living creature.

Consciousness, like the other operations of mind, functions in the life of a living creature and provides the conditions of identity of the self through the "centering" of thoughts and actions on a subject: consciousness gives the creature a sense of "care" toward what it is and what it does.

Locke thinks of consciousness as analogous to biological processes: the identity of a biological process does not depend on any kind of realizing substance, but on the persistence across time of the organism and of its organization. Likewise, the identity of consciousness does not depend on the realizing substance (whether material or immaterial) but on the persistence across time of its generating process, and like a biological process, it gives its contribution to the life of the creature.

Locke's "thinking matter hypothesis" (the thesis that purely material systems could have mental states without any immaterial soul supporting them) not only suggests, for the first time in modern philosophy, that a body could develop mental powers without the addition of a Cartesian soul, but also that the monism-dualism dilemma is the root of the conceptual problems of the traditional views of mind, because it defines "mind," "matter," and related concepts in such a way as to make impossible to understand how a subjective mental state could be part of the life of a natural creature.

Contemporary functionalism redefined these theses as follows: mind is not a thing, but a set of irreducible functional states of an organism, processes performing a causal role in the management of the organism-environment transactions. These processes are irreducible because they are "multiply realizable" in different physical systems. Multiple realizability means that since mental states are different in *type* with respect to physical states, but identical to single instances of different physical types for every single *token*, different physical systems can realize the same mental state. For this reason, mind is not reducible to any type of physical system, although mental processes mediate sensory information to produce actions: in this perspective, irreducibility does not mean that mind cannot causally act on the physical. On the contrary, mental states would be defined as causal roles or causal processes.

As the philosopher Jaegwon Kim points out (1998), functionalism has two fundamental flaws. The first is that it is compatible with contradictory theories of the mind-body relationship and especially of the ontology of consciousness. So, for example, Dennett discussed the problem of consciousness along the lines of a materialist, eliminative research project, Chalmers tried to integrate functionalism and property dualism, and Kim argued for the compatibility between func-

tionalism and type identity theory. These authors conceive of mind as a set of processes, skills, and powers and not as a “thing,” but they end up with mutually incompatible conclusions on the ontology of mind and consciousness. We could say that the old monism-dualism dilemma, which functionalism wanted to overcome, returned even in this version of the desubstantialization of mind. In the end, this view of mind is not by itself sufficient to overcome the dilemma, even though it is quite useful in characterizing the dynamic and non-reducible aspects of the mental and its interactions with the body and the environment.

The second flaw is that, during the last two decades, materialism made a strong attack against non-reductive models of mind through a version of the traditional problem of mental causation: how can we give a coherent account of causal relations involving a mental state among the *relata* under the naturalistic ontological commitment to the causal closure of the physical? Or, in another formulation, if mind is irreducible to the physical, and is therefore, in some sense, “non-physical,” we still have to explain how mind can cause something physical, such as a bodily movement.

The most sophisticated form of this problem in contemporary debate is Kim’s argument of causal exclusion. According to this argument, we have three propositions whose conjunction is inconsistent. First, physical and mental properties are different (antireductionism). Second, physics is causally closed: if we pick out a physical event and follow the chain of its causes and effects we will never find a non-physical event. Third, mental phenomena function causally.

Consider the following example: the headache (M) causes the desire (M^*) to take aspirin. Assume that the mental state M is realized by the physical state P but that it is not reducible to it and that it is causally responsible for the production of mental state M^* , in turn caused and realized by P^* . But at this point we have two possibilities.

First, we can recognize the existence of the causal relation $M \rightarrow M^*$. But this option apparently implies causal overdetermination on P^* by P and M .

Second, we can save the principle of causal closure and recognize that P is a sufficient cause of P^* . The consequence in this case is the epiphenomenalism of the level $M \rightarrow M^*$, because M would no longer be regarded as the cause of M^* .

My hypothesis is that we can overcome the problems of non-reductive models if we work out a theory of psychophysical relation that places the desubstantialization of mind within a new naturalistic framework, overcoming the monism-dualism dilemma.

This framework has its genesis in Locke’s *Essay Concerning Human Understanding* (1975) and continues in Searle’s philosophy.

The present work reconstructs Searle’s philosophy of mind. It focuses on the relationship between consciousness, mental causation, and holism (the view that mental states exist only as a part of a network of other mental states) that allows us to conceive the conscious mind as responsible for the management of the organism-environment causal relations through the development of a preconceptual, primitive sense of the self.

Engaged in current debates in philosophy of mind and cognitive science for more than twenty-five years, Searle is perhaps the most famous philosopher who recognized conscious phenomena as paradigmatic anomalies with respect to the materialist view of mind.

Within the wide philosophical literature written over the past decades on the problem of consciousness, one of the novelties of Searle's theory is that it can offer a new conception of mind and its relation to nature, avoiding the constraints deriving from the traditional monism-dualism dilemma through a radical critique of the structure of the problem.

Searle argues, on the contrary, that the reduction of theoretical options to monism and dualism is determined by what he terms "conceptual dualism," a set of theoretical assumptions implicitly shared by all the participants but rationally unjustified. Its main assumption is the so-called exclusion principle: mental and physical are mutually independent and exclusive ontological categories. But scientific theories, Searle says, show an explanatory device that can account for the existence of mind without reducing or eliminating it: mind as such, then, would be compatible with the physical as such.

This is the core thesis of "biological naturalism," as Searle calls his theory: as the states of a physical system are caused by microstructural processes and realized as macrostructural states or processes, so consciousness, with its peculiar features, is a state or process realized in the brain whose causal bases are at the lower, microstructural level of organization of the brain itself. In a sentence, conscious states are caused by and realized in the physical structure of the brain.

This relation has, according to Searle (1992, pp. 83–93), the advantage of placing the psychophysical relation within the explanatory framework of natural sciences. We can take as examples the way microphysics explains the emergence of macrofeatures such as the liquid state of water, or the way genetics and evolutionary theory explain the mechanisms underlying the development of a radically new phenomenon such as the emergence of life. These theories explain their phenomena through the same kind of relation that biological naturalism uses to explain consciousness: a system is composed of a microstructure that, at the same time, causes and realizes system macrofeatures. So, according to Searle, the conscious mind is a system, process, state, or macrofeature which causally emerges from and is realized by the brain or by any physical system with equivalent causal powers. Since consciousness is an ontologically subjective mental *and* physical process, we cannot reduce it to ontologically objective phenomena, although consciousness can be scientifically explained.

In this sense, biological naturalism is a version of the desubstantialization of mind, while we can see functionalism as an incomplete statement of it.

We could think that, as Searle has said (in conversation), functionalism shows some kind of "bashfulness" in seeing itself as engaged in an attempt to overcome monism and dualism. The reasons why I understand it this way are the denial of type identity theory and the redefinition of mind as a set of embodied processes. On my interpretation, functionalism is better viewed as an incomplete

or misleading version of the desubstantialization of mind and its flaws will emerge in some more detail in this work. Here I will only give a brief sketch of them. On one side, functionalism is still a version of materialism because it tries to define the mental in non-mental terms (that is why the label of “non-reductive physicalism” could be misleading). On the other side, functionalism is incomplete because it allows mutually contradictory theories of the ontology of mind (for example, Dennett or Kim’s materialism and Chalmers’s dualism).

Biological naturalism on the contrary can answer these flaws. It integrates the desubstantialization with the critique of conceptual dualism so that it can overcome the incompleteness of functionalism. In addition, biological naturalism is not misleading with its non-reductionism because it not only recognizes multiple realizability, but also “saves” subjectivity as an irreducible part of nature.

Searle argues against materialism for the irreducibility of the subjective and qualitative consciousness of phenomena whose ontology is a third-person one (like behavior, brain processes, and functional causal roles). He also holds, against dualism, that consciousness is a part of the biological order of the natural world, as a feature causally emergent from highly complex and developed biological systems.

Once we clarify the obsolescence and inconsistency of the philosophical vocabulary for the discussion of the problem of consciousness, the recognition of the existence and irreducibility of consciousness would not lead to Cartesian dualism, but to the transformation of the mystery of consciousness into the (scientific) problem of consciousness:

The mystery is not a metaphysical obstacle to ever understanding how the brain works; rather the sense of mystery derives from the fact that at present we not only do not know how it works, but we do not even have a clear idea of how the brain *could* work to cause consciousness . . . But we have been in similar situations before. A hundred years ago it seemed a mystery that mere matter could be *alive*. . . . The mystery was resolved not just because the mechanists won and the vitalists lost the debate, but because we got a much richer conception of the mechanisms involved. Similarly with the brain. The sense of mystery will be removed when we understand the biology of consciousness with the same depth of understanding that we now understand the biology of life. (Searle, 1997, p. 201)

This ontology of mind does not solve the scientific problem of consciousness, since biological naturalism does not tell us how the brain causes and realizes the conscious mind. Biological naturalism just describes the form that an explanation of psychophysical relation has to have. Its aim is to provide the conceptual tools for a naturalistic and non-reductive research project through the removal of the logical or cultural obstacles to the development of a scientific theory of mind and consciousness.

This naturalistic view provides the framework for Searle's analysis of intentionality, the power of some mental states (beliefs, desires, intentions, and others) to be "about" objects and states of affairs in the world. The logical structure of intentionality emerges in conscious life and we can describe this structure as a holistic Network of mental states. This Network is permeated by a Background of biologically and culturally shaped "know how" enabling the Network to determine the conditions of satisfaction of intentional states.

On this naturalistic, pragmatic, and non-reductive account, perception and action are the biologically and logically primitive elements of the Network because their contents are causally self-referential, where causal self-referentiality is the logical feature of the intentional contents of perception and intention by which they refer to the intentional state itself as caused by (perception) or as causing (action) its conditions of satisfaction.

Searle's conception of mind opposes "the passive subjectivity of the Cartesian tradition" (1992, p. 227) by pointing out a close relation between the ontology of mind and its causal efficacy within an evolutionary framework. Since consciousness is a qualitative, subjective, and unified part of the life of a biological organism, it performs an evolutionary role in the management of the organism-environment relations. Searle suggests, on the ground of clinical reports, that this role is the creative management of the behavior of the organism.

Antonio Damasio worked out a similar hypothesis in his neuroscientific research, but the hypothesis itself can find in Searle's philosophy of mind a suitable logical analysis and philosophical framework, especially with the notion of causal self-referentiality. In the interpretation of Searle's theory that I am putting forward in the present work, the relations between the ontology of mind, its holistic structure, and its causal action in a naturalistic framework have at their center the notion of the self.

In this work, I will test the capability of Searle's theory to provide solutions to the problems of consciousness and mental causation. I will proceed through a critical assessment of Searle's theory in relation with the theories of some of the most influential authors in contemporary debate, such as Dennett, Chalmers, and Kim. I will especially argue that Searle's synthesis between desubstantialization of mind and critique of conceptual dualism overcomes the endemic problems of dualism and materialism exemplified by the mentioned authors and finds support in current scientific research on consciousness.

Francesca Di Lorenzo Ajello (2001, p. 16) already pointed out that the concepts of representation, Network, Background, and causal self-referentiality as presented in Searle's work can "foreshadow an 'enactivist' picture of our cognitive and affective states," with reference to the studies of the concept of *embodiment* made by Francisco Varela, Evan Thompson, and Eleanor Rosch (1991). I will try to show that the relations between holism, ontology of mind, and mental causation (with the crucial achievement of causal self-referentiality) made explicit by Searle have *in nuce* the core elements of the recent holistic and evolu-

tionary approaches to the problem of consciousness worked out by neuroscientists such as Gerald M. Edelman, Giulio Tononi, and Damasio.

On these views, consciousness is a holistically structured process that *qua* process is irreducible to (not identifiable with) single neurons or brain places (Edelman and Tononi, 2000). Also, the unified, holistically structured field of consciousness creates a primitive, prelinguistic, and preconceptual self-reference through which an organism can “feel” that those needs and plans of action that allow him to survive, belong to it. This makes actions more flexible and creative, and gives the organism an evolutionary advantage (Damasio, 1999).

One

PROBLEMS AND THEORIES: THE MIND-BODY PROBLEM IN CONTEMPORARY PHILOSOPHICAL DEBATE

In this chapter, I will show that the process of desubstantialization of mind does not provide a sufficient framework for a coherent ontology and that this view of mind is insufficient to overcome the traditional problems of consciousness and mental causation.

I will survey contemporary dualism and functionalism, defining the problems of these positions and describing functionalism as an incomplete statement of the desubstantialization of mind in which we can still see the monism-dualism dilemma at work.

Next, I will describe John Locke's philosophy of mind as a version of the desubstantialization that explicitly tries to overcome the dilemma within a naturalized view of mental states as "powers" and "operations" performed by a creature living in the world.

Finally, I put forward an actualization of the meaning of Locke's theses through an analysis of John R. Searle's critique against "conceptual dualism," the set of theoretical presuppositions that the philosopher regards as the fundamental (and implicit) source of the traditional problems. Critique and refutation of conceptual dualism are a first, but decisive, step toward a non-reductive naturalization of mind that could integrate the desubstantialization and realize its theoretical power.

1. Explanatory Gap, Dualism, and Antireductionism

The recent debate on the mind-body problem largely concerns the existence of an *explanatory gap* between the mental and the physical. This expression designates the thesis that we lack a logically necessary connection between physical and mental concepts such that we could derive the second from the first. Sensations, thoughts, and other mental phenomena have properties the existence of which appears to be mysterious in a universe wholly constituted by such things as physical particles, force fields, and quantum phenomena.

Joseph Levine explains the idea as follows:

The idea is this. If there is nothing we can determine about C-fiber firing that explains why having one's C-fibers fire has the qualitative character it does . . . it immediately becomes imaginable that there be C-fibers firings without the feeling of pain, and vice versa. We don't have the corresponding intuition in the case of heat and the motion of molecules . . . because

whatever there is to explain about heat is explained by its being the motion of molecules. (2002, pp. 357–358)

This is not an entirely new idea, and this is true especially for the problem of consciousness. Gottfried W. Leibniz, for example, argued that even if we could make the brain large enough that we could explore it as if it were a mill or a factory, we would still not see thoughts or perceptions, only physical mechanisms that would not explain anything (2007).

Some examples from contemporary scientific research can help us to understand the problem. The development of imaging techniques such as computerized axial tomography (CAT), positron emission tomography (PET), and functional magnetic resonance (fMRI) gives us a high-resolution representation of brain functioning. So we can “directly see,” for example with a PET scan, the brain areas firing when a subject speaks, reflects on a problem, or feels an emotion.

But we still do not know why the firing of some brain area instead of another produces a conscious process.

Consider the following example, used by Giulio Tononi (2003a, pp. 16–19). We know that the brain is necessary for the existence of consciousness, but we also know that the activity of the cerebellum is not necessary to sustain consciousness. We could not be conscious without the brain, but we could remain conscious without the cerebellum. But the cerebellum is just as chemically and structurally complex as the brain, so how can we explain the different roles played by these organs in the production of consciousness?

As Antonio Damasio says (1999, pp. 18–20), by virtue of what we know about the brain, we could discover how the brain produces a self in an organism. We could understand how an organism acquires a sense of being actor and viewer of the “movie” of conscious life through the representation of itself as engaged in the interaction with the environment. But even when we could have this explanation, we would still have a problem: we would not know how the brain creates the “movie” of consciousness.

Imaging technologies and the traditional techniques to record the electric activity of the brain (such as with the electroencephalogram) allowed scientists to achieve many discoveries, but the explanatory gap still exists. We discovered a set of correlations between neural activities and mental phenomena, but we have no generally accepted theory to explain the correlations.

Whether this situation reflects the existence of some Cartesian “ontological divide” between the mental and the physical, or is just the contingent effect of the lack of a unanimously accepted neurobiological theory, is one of the main themes of the debate. The dualist tradition, from Sir Karl Raimund Popper to David J. Chalmers, supports the first position, while neuroscientists like Tononi support the second, arguing that the explanatory gap is the result of a “theoretical laziness” that would be, according to him, typical of the present neuroscientific landscape:

Theory is the Cinderella of neurosciences. . . . The brain has such a complexity that trying to understand its functioning with a net of equations as fine grained as you want is like trying to get the sea in a net for fishes. Let theoreticians and mathematicians write their equations. The brain has more trees than the jungle, more streets than a city; it is more plastic than the sand in the desert and more changing than the waves in the sea. Who would think to reduce the continuous changing of the dunes, traffic during the peak hour, or the interlace of leaves and animals in the tropical forest to a series of equations? Nowadays, neurosciences are a ferociously empirical discipline. Data are what matters and data come from experiments. Data give prestige and power and are the alpha omega of progress (my translation). (2003a, p. 25)

Notwithstanding our remarkable progress in technology and theory, the existence of conscious phenomena brings about a widespread sense of theoretical weakness, just as in the times of Leibniz. But the debate and the great amount of philosophical and scientific research on consciousness proves the renewed importance of the theme. Philosopher Colin McGinn talks about the present situation with enthusiastic words:

Recently, consciousness has leaped naked from the closet, streaking across the intellectual landscape. People are conscious—all of them! The deep, dark secret is out. You can almost hear the sigh of relief across the learned world as theorists let loose and openly acknowledge what they have repressed so long. The Nineties are to consciousness what the Sixties were to sex. (1999, p. 44)

But McGinn also argues for an extremely radical thesis: since we devoted our best efforts to the mind-body problem and the mystery still persists, he thinks that “the time has come to admit candidly that we cannot resolve the mystery” (1997, p. 529).

McGinn is not saying that consciousness emerges by virtue of an intervention of God in the natural world. He holds, instead, that the way we form concepts and theories makes the psychophysical nexus noumenal for us, human beings produced by evolution. The brain has some natural property that can explain the link between the electrochemistry of the grey matter and the colored conscious life. Then, there must be the possibility of a theoretical account that explains it. But McGinn argues that this theory would likely be so complex that we would be akin to children standing in front of Albert Einstein’s theory of relativity. Introspection, observation, and inferential reasoning on data cannot give us direct access to the psychophysical nexus. Then, he concludes that our minds are cognitively closed with respect to the solution of the mind-body problem. The noumenon will remain under the phenomenon and we will never close the explanatory gap.

Other antireductionist philosophers do not share McGinn's radical skepticism, but argue for the irreducibility of consciousness to physical or functional states without implying any dogmatic *ignoramus, ignorabimus*.

The arguments in question present different rhetorical strategies but the same logical structure. The thesis underlying contemporary materialism states that a description of mental phenomena allowing reference only to objective phenomena would give necessary and sufficient conditions to attribute mental states to a system. Once we had complete knowledge of neurophysiology we would have nothing else to know.

The antireductionist arguments show, on the contrary, that this description would be necessarily incomplete, because we can work out counterexamples: materialism does not state sufficient conditions to attribute mental states, so that it fails (Crane, 2001b; Chalmers, 1996).

Thomas Nagel argues for his thesis by imagining the ideal expert in bat physiology. This expert could give a complete reconstruction of the functioning of the echolocation system allowing these animals to have a complete orientation in the world notwithstanding their blindness. But he would not know how to answer the question, which is the title of Nagel's paper, "What Is It Like to Be a Bat?" (1981). We cannot imagine, Nagel writes, what kind of experience of the world a bat could have. We know all that needs to be known (according to the physicalistic thesis) but all the same we cannot deduce the answer to the question. Here are two passages defining the concept of "subjective character of experience" and arguing for its irreducibility:

[F]undamentally an organism has conscious mental states if and only if there is something that it is like to *be* that organism—something it is like *for* the organism. We may call this the subjective character of experience. It is not captured by all of the familiar, recently devised reductive analyses of the mental, for all of them are logically compatible with its absence. (Ibid., p. 392)

[I]f the facts of experience . . . are accessible only from one point of view, then it is a mystery how the true character of experiences could be revealed in the physical operations of the organism. The latter is a domain of objective facts par excellence—the kind that can be observed and understood from many points of view and by individuals with differing perceptual systems. There are no comparable imaginative obstacles to the acquisition of knowledge about bat neurophysiology by human scientists, and intelligent bats or Martians could learn more about the human brain than we ever will. (Ibid., p. 397)

Frank Jackson (1997; 2002) worked out another thought experiment. Imagine Mary, the ideal expert in human neurophysiology of color vision, who has complete knowledge of her specialization, but who has spent her life in an en-

tirely black and white environment. Suppose also, that we then show her a banana. Does she learn something new when she sees the color yellow? If the answer is yes, then materialism fails.

None of these authors supports some form of spiritualism or Cartesianism. Instead, they (especially Levine) just assert that, under the presuppositions of what we know about nature and of a materialistic objectivist interpretation of the methodology of sciences (Rorty, 1982; Dennett, 1982, p. 356; 1991, p. 461), we have an explanatory gap between the physical and the mental.

Any attempt to describe mind and consciousness in objective terms necessarily “leaves out” something: the subjective and qualitative aspect of mental states. This conception of the psychophysical relation can give rise to a “weak” version of dualism, called property dualism. The explanatory gap would mirror the existence of two mutually exclusive kinds of properties (mental and physical). The “metaphysical” interpretation of the explanatory gap is usually associated with Saul A. Kripke’s argument that identity statements concerning physical and mental states are not logically necessary because we could have a logically possible world in which pain, for example, is not identical with C-fiber firings (1971). The metaphysical interpretation perhaps also underlies Nagel’s double aspect theory, while Chalmers’s property dualism supports it. Nagel apparently has this kind of position in mind when he argues for a metaphysical distinction between subjective and objective aspects of experience:

Very little work has been done on the basic question . . . whether any sense can be made of experiences’ having an objective character at all. Does it make sense, in other words, to ask what my experiences are *really* like, as opposed to how they appear to me? We cannot genuinely understand the hypothesis that their nature is captured in a physical description unless we understand the more fundamental idea that they *have* an objective nature (or that the objective processes can have a subjective nature). (1981, p. 402)

Popper, for example, holds that dualism is compatible with developments in evolutionary biology and that materialism is incoherent even from a logical point of view (Popper and Eccles, 1981, pp. 23ff.; Nannini, 2004). The epistemologist also agrees with materialism about the validity of Samuel Alexander’s dictum that “real” is whatever has causal relations with the physical world. According to Popper, since mental causation is non-disputable, mind exists.

We have to conceive mind as a world (World 2) emerging from the physical (World 1) as something ontologically new. Mind can act on World 1 by way of “grasping” the “cultural objects” in World 3 (the world of cultural entities), so that mind can change the physical world through actions, projects, and ideas. But we have to specify that Popper’s dualism, which in his self-interpretation is on the same line as Cartesian dualism (or pluralism) of substances, completely abandons René Descartes’ religious inspiration.

This is a crucial difference with respect to Popper's co-author of *The Self and Its Brain*, neurobiologist Sir John Eccles: Popper holds that the conscious mind emerges from biology and physics. For him, we could not conceive of mind as a soul separated from the body.

The widespread consensus surrounding the naturalistic, scientific world-view requires us to ask what dualism is in contemporary philosophy of mind. Many materialists (mainly eliminative materialists) thought that these considerations make the dispute on the mind-body problem purely terminological. As Richard Rorty wrote:

Most discussions of the "mind-body problem" argue for reducibility or irreducibility by tacitly choosing a sense, or senses, of "reducible" favorable to their own side . . . this . . . amounts to transforming the mind-body problem into a *scholastic* issue—an issue whose outcome doesn't make a difference to anything else, one which only specialists could care about. (Rorty, 1982, pp. 323–324)

Many authoritative voices support the deflation of the mind-body problem through elimination of the subject matter for similar reasons (such as Willard van Orman Quine, Paul K. Feyerabend, Daniel C. Dennett, Paul Churchland, and Patricia Smith Churchland). But according to many dualists and antireductionists, the dispute between materialism and dualism concerns not merely terminological issues. When we talk about the mental, we are talking about ontologically new properties unrelated to the physical world in any of the ways made familiar by natural sciences. Chalmers makes this point very well:

It is sometimes said that property dualism applies to any domain in which the properties are not themselves properties invoked by physics, or directly reducible to such properties. . . . But this sort of dualism is a very weak variety. There is nothing *fundamentally* ontologically new about properties such as fitness, as they are still logically supervenient on microphysical properties. . . . By contrast, the property dualism I advocate involves fundamentally new features of the world. Because these properties are not even logically supervenient on microphysical properties, they are nonphysical in a much stronger sense. (1996, p. 125)

Notwithstanding remarkable progress in theories and technology, the explanatory gap would be still with us because we have no logical relation between the mental and the physical. In this case the explanatory gap would correspond to an ontological gap.

2. Problems with Dualism: The Problem of Mental Causation

Even though they appear to be intuitively plausible and supported by good arguments, nowadays, antireductionism and neodualism are “minority” opinions with respect to a materialist philosophical orthodoxy.

The main reason some argue against dualism is probably based on the problem of mental causation, the main test for antireductionist views of mind since the time of Descartes. Mental causation is crucial for the construction of what Wilfrid Sellars (1963) calls our “manifest image,” the picture of a subject responsible for its actions and capable of knowledge. Because of the difficulties arising in reconciling the central role of mental causation with antireductionist assumptions concerning the non-physical nature of mind in a causally closed physical world, dualism is discredited.

We can see classical examples of this difficulty in modern philosophy with Descartes, Nicolas Malebranche, and Leibniz’s solutions to their problem: since mind is an immaterial substance, how could we account for the interaction between mind and body? But Leibniz and Malebranche’s divine mediation, like the Cartesian pineal gland, are not options in a naturalistic theoretical framework. All the same, the problem came back in recent times, *mutatis mutandis*, in a nearly identical form (Heil and Mele, 1993; Peruzzi, 2004a; 2004b; Walter and Heckmann, 2003).

The concept of mental causation (a causal relation in which one or both the *relata* are mental phenomena) is a necessary tool to understand our and others’ behavior. So if I see raining outside, I will bring an umbrella with me, because I believe that when I go outside, I will get wet. Here we have an example of causation from the physical to the mental (physical and physiological events cause a perceptual experience), from the mental to the mental (a perceptual experience causes the belief that it is raining outside, that in turn causes, with other mental states, the belief that if I go outside I will get wet), and from the mental to the physical (I take an umbrella because I intend to go outside, because I believe that it is raining, and because I believe that if I go outside I will get wet).

Examples like this explain why many philosophers think that mental causation is a matter of fact—that philosophical reflections and scientific theories must take for granted. A good example of this stance comes from Jerry Fodor’s famous passage:

I’m not really convinced that it matters very much whether the mental is physical; still less that it matters very much whether we can prove that it is. Whereas, if it isn’t literally true that my wanting is causally responsible for my reaching, and my itching is causally responsible for my scratching, and my believing is causally responsible for my saying . . . , if none of that is literally true, then practically everything I believe about anything is false and it’s the end of the world. (1990, p. 156)

According to Fodor, we not only have to take mental causation for granted, but also we must consider its existence independent of our background theory of the mind-body relation (see also Burge, 1993, pp. 102–103, 116).

Fodor's opinion is far from being unanimously accepted. As Jaegwon Kim says (1998, pp. 61–63), if we hold that mental causation exists, then we have to give an analysis of it. Mental causation has to be coherent with the assumptions of a naturalistic conceptual framework, so that the analysis has to be included in the more general debate on the mind-body problem. We can say that, if Fodor and Tyler Burge are right, then this does not exempt philosophy from reflection on, and the search for, a justification of mental causation. We have to say how a non-physical mind can act in a physical world. To see why, let us consider some examples.

Blindsight patients are partially lacking of phenomenally realized sight, but they will reliably report whether in the blind part of their visual field stands an *X* or an *O*, so that their cognitive performance (in the experimental setting) does not appear to be influenced by the lacking phenomenal realization of the information carried out by the neuronal substrate (Weiskrantz, 1986; 1988; Bornstein and Pittman, 1992; Gazzaniga, 1998).

Also, many texts report cases of epilepsy where the patients continue to engage in complex behavior (such as walking or opening a door) and routine activities (such as driving a car or playing piano) even though their conscious state is suspended during seizures (Damasio, 1999).

The mentioned examples, in which apparently the lacking of mental states does not affect the subjects' "physical" performances, raise a more general problem. Our behavior and bodily movements are subject to complete "physical explanations" that do not mention mental states.

As Michael Tye writes:

Given that there is a complete physical explanation of why your arm moved as it did, there is simply not room for a nonphysical cause to operate. . . . Subjectively felt qualities can make a difference only if they are realized or generated in a nonmagical way by underlying objective physical properties, and hence only if they are themselves physical (Tye, 1995, p. 57)

Tye gives us the link between the mind-body problem and the problem of mental causation: non-reductive theories would interpret mental causes as non-physical. The alternative is reductionism: mental states would be nothing but physical states. This is the reason why the problem does not arise for the identity theorists or eliminativists, but only for those holding that our mental aspect is irreducible to the physical (Crane, 1995, p. 211).

Another argument, called "the zombie argument," points out this connection between non-reductive or dualistic theories of mind and epiphenomenalism.

Chalmers writes:

The basic shape of the argument is clear: if it is possible to subtract the phenomenal from our world and still retain a causally closed world Z, then everything that happens in Z has a causal explanation that is independent of the phenomenal. But everything that happens in Z also happens in our world, so the causal explanation that applies in Z applies equally here. So the phenomenal is causally irrelevant. Even if conscious experience were absent, the behavior might have been caused in exactly the same way. (1996, p. 150)

Chalmers believes that mind is grounded on the physical via natural supervenience (through contingent psychophysical bridge laws). Nonetheless his view of the conscious mind makes it explanatorily irrelevant or epiphenomenal:

We can give explanations of behavior in purely physical or computational terms, terms that neither involve nor imply phenomenology. If experience is tied in some intimate way to causation, it is in a way these explanations can abstract away from. (Ibid., p. 156)

Some antireductionists (Jackson, 2002) draw on the following consequence from this argument: If the zombie argument is correct, then it shows that a physical system lacking mental states can have any kind of causal relations with the world and these relations would be logically independent of the existence of mental states. The result is epiphenomenalism. But the problem of mental causation can be present even in functionalism, the contemporary orthodox variety of materialism.

3. Materialism, Functionalism, and the Desubstantialization of Mind

Contemporary materialism expresses a wide range of positions, such as behaviorism, eliminativism, type and token identity theories, and functionalism. As Searle says, materialism is not easy to define, since even though we could say that “it would consist in the view that the microstructure of the world is entirely made up of physical particles,” we would have to take into account that this definition would be “consistent with just about any philosophy of mind, except possibly the Cartesian view” (Searle, 1992, p. 27).

Perhaps we could say that materialism is inspired by what John McDowell (1994, pp. 78ff.) would define as a “bald naturalism”: working out a theory of psychophysical relation means analyzing mental phenomena and their properties in objective, non-mental terms. Under this respect the difference between reductionism and eliminativism would reduce to a difference in rhetorical strategy:

The difference between “eliminative materialism” . . . and a “reductive” materialism that takes on the burden of identifying the problematic item in

terms of the foundational materialistic theory is . . . often best seen not so much as a doctrinal issue as a tactical issue: how might we most gracefully or effectively enlighten the confused in this instance? (Dennett, 2002, p. 244)

So behaviorists said that mental phenomena are behavior and dispositions to behavior; identity theorists said that mental phenomena are neural states; functionalists said that mental phenomena are causal roles; and computationalists said that these roles are applications of algorithmic procedures of information processing that mediate sensory inputs and behavioral outputs. All of these forms of materialism have, in this sense, a reductionist aim: the complete analysis of mind in terms of the non-mental. Fodor, for example, well exemplifies his research program with this sentence on intentionality: "If aboutness is real it must really be something else." (Fodor, 1987, p. 97)

If on one side, functionalism tries to reduce mind to objective causal roles, we have to admit that, for reasons that I am going to explain, it tries to outstrip the boundaries of the monism-dualism dilemma. To say why, I need to analyze this position in some detail.

Functionalism emerged between the 1960s and the 1970s as the materialist orthodox position in philosophy of mind and as the theoretical framework of research projects in cognitive science. If on one side, the success of this perspective within the newborn cognitive psychology is due to the fruitful exchange of philosophy, psychology, computer science, artificial intelligence (A.I.), and psychology, and to the fall of behaviorism produced by Noam Chomsky's linguistics, on the other side, functionalism promised to overcome the conceptual problems of materialism and dualism.

Sir Hilary Putnam is the founder of contemporary functionalism in its computational version. He wrote a series of papers, collected in his *Mind, Language, and Reality* (1975), in which the new theoretical perspective is presented as an alternative to behaviorism and identity theory with the arguments of the "super-spartan" and of multiple realizability.

Here is his analysis of pain, grounded in the hypothesis that pain is not a physical-chemical state of the brain, but a functional state of the entire organism:

- (1) All organisms capable of feeling pain are Probabilistic Automata.
- (2) Every organism capable of feeling pain possesses at least one Description of a certain kind. (Being able of feeling pain is possessing an appropriate kind of Functional Organization.)
- (3) No organism capable of feeling pain possesses a decomposition into parts which separately possess Descriptions of the kind referred to in (2).
- (4) For every Description of the kind referred to in (2), there exists a subset of the sensory inputs such that an organism with that Description is in pain when and only when some of its sensory inputs are in that subset. (Putnam, 1975c, p. 434)

The key thesis is the second: We can analyze the nature of pain with the formal (without reference to the particular physical realization) description of causal relations mediating sensory inputs (for example pain from a wound) and behavioral outputs (for example, a scream).

The core thesis of functionalism is the contemporary version of a philosophical position whose authors tried to work out a third view between monism and dualism to account for the peculiarity of mental phenomena and their embodiment as a part of the natural world. Examples of this position are philosophers such as Aristotle, Locke, William James, Friedrich Nietzsche, and Ernst Mach (Rosciglione, 2005; Guli, 2007).

This thesis, which I call desubstantialization of mind, holds that mind and mental states are not things or faculties, but processes, abilities, and powers of an organism, which are dependent on the physical dimension of the organism (its body in interaction with the environment) but that, *qua* processes, are irreducible to the physical itself. Mind is a causally real level of the functioning of an organism, although it has peculiar properties differentiating it from ordinary physical and biological phenomena.

In its functionalistic version, the desubstantialization of mind says that mental states are functional states (causal processes) of an organism implemented by the brain and responsible for the mediation between sensory inputs and behavioral outputs. As processes, mind and mental states are not identifiable with any kind of physical substratum. So, to preserve the identity and existence of the process is not necessarily a specific type of realizing substance, but the right kind of input-output relations. This means that mental processes are multiply realizable. They are supportable in principle by any kind of substratum stable enough to allow the process to carry out all the steps articulating the input-output mediation. In Dennett's words:

Functionalism is the idea that handsome is as handsome does. . . . The actual matter of the brain . . . is roughly . . . irrelevant. . . . According to this tempting proposal, even the underlying micro architecture of the brain's connection can be ignored for many purposes . . . since it has been proven by computer scientists that any function that can be computed by one specific computational architecture can also be computed (perhaps much less efficiently) by another architecture. If all that matters is computation, we can ignore the brain's wiring diagram, and its chemistry, and just worry about the "software" that runs on it. (2005, pp. 16–17)

Multiple realizability provides functionalism with a minimal materialism according to which, although a mental type *M* is not reducible to a physical type *P*, single tokens of *M*; $m_1, m_2 \dots m_n$ are identical to single tokens of many *P* types. Multiple realizability is, therefore, compatible with token identity theory.

Unlike behaviorism, functionalism can conceive of mental states as internal devices responsible for the behavior of the organism or system; also, unlike type

identity theory, it can state that mind has some degree of autonomy with respect to the implementing physical system by virtue of its multiple realizability.

4. Problems with Functionalism: Consciousness, Mental Causation, and Incompleteness

Early functionalism (or Putnam's style functionalism) put forward a general view of mind aimed at analyzing conscious and unconscious mental states. But the antireductionist arguments claim that we cannot functionally analyze conscious states. Examples of these arguments are those put forward by Nagel and Jackson, and the argument of absent qualia, also known as the Chinese Nation Argument (a variation of the zombie argument), worked out by Ned Block.

Putnam (1981, p. 81) eventually gave up the task of giving a functionalist analysis of consciousness. He first returned to the old identity theory for qualia (another philosophical term for conscious states) and then (1988) abandoned functionalism, with an extension of the multiple realizability argument against functionalism: A given mental state is not identical with a given functional state in that mental states are multiply realizable on different functional states.

The aim of Block's argument (1978) is to show that functionalism, like behaviorism, does not state sufficient (or sufficiently compelling) conditions for the existence of mental states because it attributes them to systems lacking of mental states. According to this argument, nothing in functional analyses would prohibit giving the entire Chinese population some mutually linked radio transmitters, so that the individuals could interact and reproduce, for example, the formal input-output description of a pain, or even of a belief or desire. We can also imagine that the system of radio transmitters is linked to a robotic body performing the behavioral outputs resulting from the interactions of the Chinese population.

Is the system of radio transmitters sufficient to consider the Chinese population as a single subject with its mental states? Intuitively we would answer no, and if the argument works, the consequence is that identity of functional organization and identity of performance do not guarantee identity of mental states.

In many respects, Searle's Chinese room argument (1981; 1984) has the same logical structure of the arguments worked out by Jackson, Nagel, and Block. His argument points out the insufficiency of the postulate that performance identity implies mental states identity, a principle that works deeply in the computational version of functionalism underlying the research in the field of A. I. The claim of what Searle calls "strong" A. I. is that a properly programmed physical system does not just simulate a mind; instead, it is a mind.

We can formally state Searle's argument in three premises and a conclusion: programs are entirely syntactical; minds have semantics; syntax is not the same, or by itself sufficient for, semantics. Therefore, programs are not minds (Searle, 1997, pp. 11–12).

The first premise states the formal or purely syntactical nature of computation, namely the abstract character of mental processes with respect to their par-

ticular physical realization. The second premise captures a fundamental feature of intentional mental states: They have a semantic content; they refer to objects and states of affairs in the world. The third premise concerns the relationships between semantics and syntax. No formal system determines by itself its semantic interpretation. Finally, the conclusion refutes strong A. I.: Since programs are syntactic entities and minds have semantic contents, programs are not minds because syntax is not sufficient to determine semantics.

The Chinese room thought experiment shows the validity of the third premise and of the conclusion:

Imagine that you carry out the steps in a program for answering questions in a language you do not understand. I do not understand Chinese, so I imagine that I am locked in a room with a lot of boxes of Chinese symbols (the database), I get small bunches of Chinese symbols passed to me (questions in Chinese), and I look up in a rule book (the program) what I am supposed to do. I perform certain operations on the symbols in accordance with the rules (that is, I carry out the steps in the program) and give back small bunches of symbols (answers to the questions) to those outside the room. I am the computer implementing a program for answering questions in Chinese, but all the same I do not understand a word of Chinese. And this is the point: *if I do not understand Chinese solely on the basis of implementing a computer program for understanding Chinese, then neither does any other digital computer solely on that basis, because no digital computer has anything I do not have.* (Ibid., p. 8)

We can say, against Douglas Hofstadter (1980; 1981) and Richard Rorty (1980) that the argument and the conclusion do not imply that machines cannot think without the addition of some Cartesian style “extra ingredient.” Brains, Searle says, are organic machines and they can think, so this interpretation cannot be right (1997, p. 202).

We also know that everything in the universe has a description in purely syntactic-formal terms; consequently, a formal description could also simulate brain functioning (Searle, 1992, pp. 200–201, 208–209). But—and here is the difference between strong and weak A. I.—simulating is not the same as duplicating, unless we do not want to postulate the identity of behavioral performance and mental states, returning to behaviorism.

We could say, based on these arguments, that the desubstantialization of mind in its functionalistic version is not sufficient as an analysis of mental phenomena and that, as a consequence of the debate that followed these arguments, it became compatible with mutually contradictory solutions of the problem of consciousness.

Dennett and Chalmers, for example, share the functionalistic thesis that mind is a multiply realizable process but, as far as consciousness is concerned, they derive from the same premise a form of eliminative monism and a property dualism respectively. Also, Kim puts forward a version of functionalism that makes it

compatible with a form of type identity theory and, therefore, with a strong reductionism identifying mental processes and their causal powers with their respective physical realizers and their properties.

If the desubstantialization of mind is compatible with the whole range of theoretical options available within a naturalistic framework, then it is insufficient to define a coherent theory of the ontology of mind. As Kim said, philosophers of mind accepted the dissolution of type identity theory without any return to dualism because with different modalities, Donald Davidson's anomalous monism (the theory that mental states are tokens of physical states, although we have no psychophysical bridge laws) and Putnam's antireductionist functionalism appeared to synthesize physicalist and antireductionist claims with the theses of multiple realizability and token identity (Kim, 1998, pp. 2–3).

The return to type identity theory put forward by Kim is motivated by the fact that a non-reductive interpretation of functionalism would inherit from dualism the inability to give a coherent account of mental causation. The thesis sounds *prima facie* not reliable because, according to functionalism, mental states are causal roles. But if we look more closely at functionalist proposals in detail, then we can shed light on the problem.

To begin with, we can apply the zombie argument to functionalism. Functionalism states a thesis concerning the nature of mental states. It defines them as “nothing but” the set of causal roles that a physical state plays in the production of the manifest behavior of an organism through the mediation between sensory inputs and behavioral outputs. If we have the correct formal pattern of input-output causal relations, we have all we need to define mental states: a necessary and sufficient connection between having mental states and the proper behavioral pattern.

Next, we can coherently conceive of a case in which we can satisfy a functionalist analysis but the system does not have mental states. Conceiving of a being behaving exactly like a mindful agent but lacking mental states is logically possible (non self-contradictory) if we accept arguments like Nagel's, Jackson's, Block's, and Searle's. Then functionalism fails because the causal-behavioral analysis does not state sufficient conditions for the existence of mental states.

However, the most powerful challenge against functionalism on the problem of mental causation comes from Kim's causal exclusion dilemma. Consider what Kim writes about a thesis put forward by Block:

Being in pain, on the functionalist account, is the second-order property of instantiating some property or other, presumably a neurobiological property that causes wincings, cryings, limb withdrawals, and the like. If so, can pain itself, in addition to its neural realizers, be considered a cause of wincings and cryings? (Ibid., p. 51)

Some physical systems exemplify the second-order property of exemplifying input-output patterns (the processes situated, for example, between hurt and scream) realized by first-order physical properties of the system. The question is

this: Do the second-order properties have causal relevance by themselves to explain the behavior of the system? What is it that brings about the scream? Is it the pain or the neuron firings?

Let us consider the following analogy (put forward by Block): A torero's cape has the property to be red and the property to be red has, in turn, the property to provoke bulls. Does the provocativeness of red do some further causal work not already carried out by the redness? Here is Kim's answer:

The answer obviously is none: given the color of the cape as a full cause, there is no *additional* causal work left for its provocativeness, or anything else. It should be clear then that there is a real problem, the exclusion problem, in recognizing second-order properties as causally efficacious in addition to their realizers (Ibid., p. 53)

The second-order property of feeling pain does not carry out any role that is not already performed by the first-order properties of the physical realizers: whether pain is realized does not make any difference under the aspect of causal relevance. The work is already done by neuron firings. The reason is that if functional properties are irreducible second-order properties, then their causal powers fall under the more general problem of causal exclusion.

Kim argues that this holds equally for neodualism and functionalism. The problem formalizes the intuition supporting the zombie argument and the supposed non-physical nature of consciousness, but here Kim makes a further point: Minimal physicalism (token identity theory) does not solve the problem because it prevents us to reduce mental properties to physical properties. Here is Kim's formulation:

Suppose . . . mental event m , occurring at time t , causes physical event p , and let us suppose that this causal relation holds in virtue of the fact that m is an event of mental kind M and p an event of physical kind P . Does p also have a physical cause at t , an event of some kind N ?

To acknowledge mental event m (occurring at t) as a cause of physical event p but denying that p has a physical cause at t would be a clear violation of the causal closure of the physical domain. . . . But to acknowledge that p also has a physical cause p^* , at t is to invite the question: Given that p has a physical cause p^* , what causal work is left for m to contribute? The physical cause therefore threatens to exclude, and pre-empt, the mental cause. This is the problem of causal exclusion. The antireductive physicalist who wants to remain a mental realist, therefore, must give an account of how the mental cause and the physical cause of one and the same event are related to each other. Token physicalism . . . is not enough, since the question ultimately involves the causal efficacy of mental *properties*, and antireductionism precludes their reductive identification with physical properties. Thus the problem of causal exclusion is to answer this question: *Given that*

every physical event that has a cause has a physical cause, how is a mental cause also possible? (Ibid., pp. 37–38)

I will deepen the problem and the argument later in this work. For now, suffice it to say that if Kim is right, then the problem of mental causation arises even for functionalism because it concerns the compatibility of naturalism and antireductionism in general. Apparently we have to choose between the naturalistic commitment to the causal closure of the physical (physical events have only physical causes and effects) and the reality of mental causation under a non-reductive interpretation of the ontology of mind. We cannot have both.

If the analysis I put forward is correct, we could understand why the desubstantialization of mind made by contemporary functionalism tends to fall under one of the horns of that monism-dualism dilemma it originally aimed to overcome.

Roughly speaking, we could say that the materialist part of functionalism, its analysis of mental phenomena in the objective terminology of causal roles and token identity theory, exposes this conception of mind to the antireductionist arguments. On the other hand, the non-reductive part of functionalism, grounded in the metaphysical thesis of multiple realizability or in Davidson's (1980b) or Fodor's (2002) epistemic arguments exposes the thesis to the challenge of working out a conception of mental causation which has to be coherent with the principle of causal closure.

Under this framework, the interpretation of the nature of mind in eliminative (Dennett), neodualist (Chalmers), or identitist (Kim) terms would not appear arbitrary, since the logical space of the thesis allows for all of the mentioned theoretical options. So, if we want to preserve an objective conception of science grounded in a strong naturalism, we can withdraw on the materialist part of functionalism and deny any validity to the antireductionist arguments, trying to reduce or eliminate the subjective and qualitative dimension of mind.

Alternatively, we could withdraw on the non-reductive part of the thesis. In this case, we will have to give a reliable account of mental causation to reconcile the "novelty" of the conscious mind with its embodiment in the natural world.

Conceiving of mind not as a thing separated from the rest of the natural world, but as a set of dynamic processes and abilities responsible for the interaction of the system with its environment, is a remarkable conceptual progress with respect to the modern disputes on what kind of substance the mind is. But the old monism-dualism dilemma appears to return even in the new theoretical framework.

To understand whether those problems are internal to the desubstantialization of mind as such or they depend on the particular contemporary and functionalistic version of it, we have to take a step back and reconstruct one of the most meaningful genetic moments of it through an analysis of Locke's theses in the *Essay Concerning Human Understanding* (1975), where we can perhaps find a version of the desubstantialization of mind consistent with a deflation of the monism-dualism dilemma.

5. Return to the Roots of the Desubstantialization of Mind: The Mind-Body Problem in John Locke's Philosophy

The aim of this section is to reconstruct the concepts and arguments determining the paradigmatic antisubstantialist turn in John Locke's *Essay Concerning Human Understanding*. I will also test my interpretive hypothesis that in the *Essay*, Locke argues for the desubstantialization of mind and for the necessity of a refutation of the monism-dualism dilemma with a deflation of the traditional mind-body problem.

Descartes' conception of a passive subject, solipsistically opposed to the world (a view happily labeled by contemporary neopragmatists as "the Myth of the Given") depends on a conception of knowledge as mere mirroring of reality and Locke is usually seen as a part of Descartes' tradition. This interpretive pattern finds in Rorty's *Philosophy and the Mirror of Nature* (1979) its most efficacious presentation in the analytic landscape, where the philosopher reads Locke's philosophy as a genetic moment of the foundationalist epistemological tradition (see especially chap. 3). But the neopragmatist historiography is progressively abandoning this reading. Also, Rorty does not make use of textual support to argue his thesis. He just quotes critiques to Locke's approach to knowledge put forward by philosophers like Sellars. According to Yasuhiko Tomida, for example, we should instead see Locke as a natural ally against a Cartesian view of ourselves (2001, p. 197).

We could observe that the problematic assimilation of Locke's philosophy to Rorty's reading depends first on the presence in the *Essay* of elements that give value to a pragmatist and contextualist interpretation of knowledge, where we could see knowledge as a function performed by a living creature and devoted to reach well-being in this world. So if by contextualism we mean the idea that knowledge does not have "foundations," since it is justified by the set of cultural productions and social practices in a particular historical context, then the famous analogy between the philosopher and the underlaborer is a clue of a possible alternative interpretation with respect to Rorty's:

The Commonwealth of Learning, is not at this time without Master-Builders, whose mighty Design, in advancing the Sciences, will leave lasting Monuments to the Admiration of Posterity; But every one must not hope to be a Boyle, or a Sydenham; and in a Age that produces such Masters, as the Great Huygenius, and the incomparable Mr. Newton. . . . 'tis Ambition enough to be employed as an Under-Labourer in clearing Ground a little, and removing some of the Rubbish, that lies in the way to Knowledge; which certainly had been much more advanced in the World, if the Endeavours of ingenious and industrious Men had not been much cumbred with the learned but frivolous use of uncouth, affected, or unintelligible Terms, introduced into the Sciences, and there made an Art of, to that De-

gree, that Philosophy, which is nothing but the true Knowledge of Things, was thought unfit, or incapable to be brought into well-bred Company, and polite Conversation. (Locke, 1975, pp. 9–10)

We could say that where Descartes (1993; 1998) describes his picture of mind on the grounds of a “mirroring” conception of knowledge, Locke grounds a pragmatic and instrumentalist view in the naturalistic character of his conception of human beings as creatures living in the world whose knowledge has to be assessed in relation to its power of efficaciously conducting actions.

This point emerges, for example, from the Introduction to the *Essay*, where Locke writes that the aim of an inquiry on knowledge has to be to “find out those Measures, whereby a rational Creature put in that state, which Man is in, in this World, may, and ought to govern his Opinions, and Actions depending thereon” (ibid., 1.1.6). On this basis, he then criticizes the rationalistic and skeptical views of knowledge, claiming that they share the false premise that knowledge has nothing to do with the conduction of action and with the reaching of well-being in this world:

We shall not have much Reason to complain the narrowness of our Minds, if we will but employ them about what may be of use to us; for of that they are very capable: And it will be an unpardonable, as well as Childish Peevishness, if we undervalue the Advantages of our Knowledge, and neglect to improve it to the ends for which it was given us, because there are some things that are set out of the reach of it. It will be no Excuse to an idle and untoward Servant, who would not attend his Business by Candle-light, to plead that he had not broad Sun-shine. The Candle, that is set up in us, shines bright enough for all our Purposes. . . . If we will disbelieve every thing, because we cannot certainly know all things; we shall do much—what as wisely as he, who would not use his Legs, but sit still and perish, because he had no Wings to fly. (Ibid., 1.1.5)

This view of subjectivity and knowledge is a *leit-motiv* of Locke’s philosophy (Viano, 1960; Gargani, 1966; Tagliagambe, 1980) that provides the framework of Locke’s analysis of mind and of its relations with the rest of the natural world, which he developed in a non-systematic way.

Locke’s analysis first takes the form of a denunciation of the explanatory emptiness of the substantialist metaphysics underlying Cartesian dualism. Then he tries to substitute these empty views with a conception of mind grounded in the concepts of “power,” “action,” and “operation” of an agent.

Locke criticizes the tautological character of the substantialist metaphysics (Locke, 1975, 2.23) and applied his objections to the analysis of mind, pointing out that a substantialist conception of thinking, sensibility, cognition, and volition can stand only on the basis of a question begging argument.

In the second part of his critique, Locke reconstructs the categorical mistake

that we make when we treat a set of powers as distinct “agents” or “faculties”:

[T]he fault has been, that Faculties has been spoken of, and represented, as so many distinct Agents. For it being asked, what it was that digested the Meat in our Stomachs? It was a ready, and very satisfactory Answer, to say, That it was the *digestive Faculty*. What was it that made any thing come out of the Body? *The expulsive Faculty*. What moved? *The Motive Faculty*. And so in the Mind, the *intellectual Faculty*, or Understanding, understood; and the *elective Faculty*, or the Will, willed or commanded: which is in short to say, That the ability to digest digested, and the ability to move, moved; and the ability to understand, understood. For *Faculty*, *Ability*, and *Power*, I think, are but different names of the same things: Which ways of speaking, when put into more intelligible Words, will, I think, amount to this much; That Digestion is performed by something that is able to digest; Motion by something able to move; and Understanding by something able to understand. (Ibid., 2.21.20)

Also, Locke points out the question-begging character underlying the picture of thinking as the “essence” of mind that commits its supporter to the implausible thesis that “soul always thinks”:

therefore, though thinking be supposed never so much the proper Action of the Soul; yet it is not necessary, to suppose, that it should be always thinking, always in Action. . . . 'Tis doubted whether I thought all last night, or no; the Question being about a matter of fact, 'tis begging it, to bring, as a proof for it, an Hypothesis, which is the very thing in dispute: by which way one may prove any thing. . . . But he, that would not deceive himself, ought to build his Hypothesis on matter of fact, and make it out by sensible experience, and not presume on matter of fact, because of his Hypothesis, that is, because he supposes it to be so: which way of proving, amounts to this, That I must necessarily think all last night, because another supposes I always think, though I myself cannot perceive, that I always do so. (Ibid., 2.1.10)

Against this logically invalid and explanatorily empty picture, Locke develops his antisubstantialist and antiCartesian view, that the philosopher makes explicit not only with his famous view of thought as “what motion is to the Body, not its Essence, but one of its Operations” (ibid.), but also with the view of mental states as “powers” performed not by a set of “things,” but by an agent, since “it is the Mind that operates, and exerts these Powers; it is the Man that does the Action; it is the Agent that has the Power, or is able to do. For Powers are Relations, not Agents” (ibid., 2.21.19).

Locke applies this relational view of mental states to consciousness, which he construes as analogous with the biological processes of a living creature, whose identity does not depend on the persistence of the realizing matter, but on

the continuity of the processes across time. Consider the example of an oak tree. Its identity does not depend on the identity of the mass of matter composing it, but on “an Organization of those Parts, as is fit to receive, and distribute nourishment, so as to continue, and frame the Wood, Bark, and Leaves” (ibid., 2.27.4).

Also noteworthy is Locke’s idea that having consciousness necessarily implies the development of a form of self-care, a form of worry for the consequences of actions and thoughts as they are felt as actions and thoughts belonging to an individual subject. If this interpretation is correct, in passages like the following, we could find the genesis of a line of thought that in contemporary philosophy and science found new expressions, for example with the theories worked out by Searle and Damasio on the causal and evolutionary role of consciousness. In these new naturalistic frameworks, this aspect of consciousness plays a crucial role in the management of the mind-world relationship. Here is what Locke says:

Person, as I take it, is the name for this *self*. Where-ever a Man finds, what he calls *himself*, there I think another may say is the same *Person*. It is a Forensick Term appropriating Actions and their Merit; and so belongs only to intelligent Agents capable of Law, and Happiness and Misery. This personality extends it *self* beyond present Existence to what is past, only by consciousness, whereby it becomes concerned and accountable, owns and imputes to it *self* past Actions, just upon the same ground, and for the same reason, that it does the present. All which is founded in a concern for Happiness the unavoidable concomitant of consciousness, that which is conscious of Pleasure and Pain, desiring, that that *self*, that is conscious, should be happy. And therefore whatever past Actions it cannot reconcile or appropriate to that present *self* by consciousness, it can be no more concerned in, than if they had never been done: And to receive Pleasure or Pain; i.e. Reward or Punishment, on the account of any such Action, is all one, as to be made happy or miserable in its first being, without any demerit at all. (Ibid., 2. 27.26)

But talking about the “faculties of the body” and the “faculties of the mind” apparently means recognizing some kind of dualism between the mental and the physical. Does Locke believe in a metaphysical distinction between them?

A great part of the literature in the last three decades saw Locke as a dualist (Odegard, 1970; Bennett, 1999). This dualist interpretation is a shift with respect to the original attack made against Locke as a materialist, that was formulated not only by Leibniz, but also by politically and culturally influential people such as theologian Edward Stillingfleet, Bishop of Worcester (1689). Locke’s critics argued his supposed materialism based on his thinking matter hypothesis and the analysis of personal identity, whose terms were independent from the existence of the soul (McCann, 1999). In addition, as Lynn Holt (1999) said, at the

time of the debate opened by the publication of the *Essay*, the materialistic interpretation was widespread at each social level.

However, some texts apparently commit Locke to a dualistic conception. For example, consider the following passage:

It is for want of reflection, that we are apt to think, that our Senses shew us nothing but material things. Every act of sensation, when duly considered, gives us an equal view of both parts of nature, the Corporeal and Spiritual. For whilst I know, by seeing or hearing, *etc.* that there is some Corporeal Being without me, the Object of that sensation, I do more certainly know, that there is some Spiritual Being within me, that sees and hears. This I must be convinced cannot be the action of bare insensible matter; nor ever could be without an immaterial thinking Being. (Locke, 1975, 2.23.15)

Here Locke says that the existence of mental phenomena is such that “I must be convinced” of the existence of an immaterial substance which is their bearer. Recent critical literature pointed out that this and other passages need a wider context to account for Locke’s naturalistic conception of mind and his nominalism, the theory of universals put forward in Book Three of the *Essay* according to which universal terms refer to general ideas in the speakers’ minds and not to real essences (Jolley, 1999, pp. 84ff.).

The interpretive thesis holds that mind and body are general terms and as such, on the basis of nominalism, they do not name the ontological structure of reality; instead, they are interpretive frameworks created by human beings to give them back a coherent picture of a world, which is irreducibly plural and in which human beings have to be capable of life, orientation, and action (Locke, 1975, 3.3).

Universals, “Creatures of the Understanding” (ibid., 3.3.11), reflect powers and limits of understanding and also the competence of the people who contributed to the formation of our concepts. The concepts of mind and body have the same features of the other general ideas and this could be seen in the restrictions posed by Locke to the assertions apparently interpretable in a dualistic manner. So, for example, Locke writes, “I must be convinced” of the existence of an immaterial being within me, because “nor apprehending how they [the operations of mind] can belong to Body, or be produced by it, we are apt to think these the Actions of some other *Substance*, which we call *Spirit*” (ibid., 2.23.5).

In both cases, Locke is putting some psychological restrictions on the application of the supposed dualism, pointing out our limited epistemic situation.

So a possible alternative interpretation would hold that mind and body are abstract concepts created by human beings to name what Locke calls “the two parts of nature” (the spiritual and the corporeal) where the first is analyzed in terms of relations, operations, and powers performed by living creatures. But to say that these powers are metaphysically independent of the body, or radically incompatible with it, is not a well justified claim because this belief could be grounded only in our ignorance and in the way we create and use the concepts of

mind and body. We can also remember, for example, that Locke criticized Descartes' thesis about the impossibility to attribute spatial features to minds (ibid., 2.23.19–21). According to Locke, in light of the little we know, we are not allowed to understand how these operations could be produced by the body. But this could be a historical accident.

Mind is different from the body, which is defined as mass and movement, but deriving dualistic conclusions from these premises is not a valid inference unless we have some independent argument proving the metaphysical incompatibility of mind and body.

Two fundamental reasons support this interpretation. For Locke, consciousness is a process that gives a more efficacious orientation in the natural world by virtue of its providing a sense of "ownership" of thoughts, actions, and their consequences. Consciousness creates self-care, the project of future actions, and the assumption of responsibility. This naturalistic, pragmatic view of consciousness hardly fits the usual patterns of dualism and materialism since Locke does not want to eliminate the natural and the subjective character of mind. Instead, he sees the subjective mind as linked to the needs of a natural creature. This leads to the second reason; cognition and volition are linked to this self-concern.

Locke sees cognitive and extra-cognitive mental states as closely linked, since they allow a living creature to conduct its actions in an efficacious way. Pleasure and pain, for example, provide us with a concern that drives and motivates our thinking and acting:

[God] having also given a power to our Minds, in several instances, to chuse, amongst its *Ideas*, which it will think on, and to pursue the enquiry of this or that Subject with consideration and attention, to excite us to these Actions of thinking and motion, that we are capable of, has been pleased to join to several Thoughts, and several Sensations, a *perception of Delight*. If this were wholly separated from all our outward Sensations, and inward Thoughts, we should have no reason to prefer one Thought or Action, to another; Negligence, to Attention; or Motion, to Rest. And so we should neither stir our Bodies, nor employ our Minds; but let our Thoughts (if I may so call it) run a drift, without any direction or design; and suffer the *Ideas* of our Minds, like unregarded shadows, to make their appearance there, as it happen'd, without attending to them. In which state Man, however furnished with the Faculties of Understanding and Will, would be a very idle unactive creature and pass his time only in a lazy lethargick Dream. (Ibid., 2.7.3)

If this analysis is correct, then Locke's philosophy of mind is hard to classify within one of the traditional monist or dualist options because he recognizes the peculiar features that make mind different from the rest of the natural world. But he also points out that the function of mind is to serve the needs of a natural creature. Mind exists and is a part of nature. It is a set of natural, subjective processes the

function of which is the good conduction of actions for the benefit of living beings. As Edwin McCann writes:

Locke is a particularly interesting case. He was accused of being a materialist or at least a fellow-traveler of materialism by Leibniz and by many contemporary critics in England; more recently, he has been seen as a dualist. He was neither; as usual, he preferred to remain agnostic in an area where human reason could make no reliable pronouncements. (1995, p. 343)

McCann's conclusion appears sound, although the adduced reason (Locke's skepticism) is slightly different from the reading I am offering. I would interpret Locke's agnosticism in terms of the deflation of the Cartesian mind-body problem, on the grounds that the metaphysical categories of mind and body used to state the problem are just tools used by human beings. If this analysis is correct, then the dilemmatic formulation of the mind-body problem is wrong because we have no compelling reason to think that mind cannot be, as a matter of definition, physical, or that a material system cannot develop the power of thinking by means of the proper organization of its part and not with the addition of some non-natural element (something like a Cartesian *res cogitans*).

We can interpret the thinking matter hypothesis as an argument supporting the mentioned conclusion:

We have the *Ideas* of *Matter* and *Thinking*, but possibly shall never be able to know, whether any mere material Being thinks, or no; it being impossible for us, by the contemplation of our own *Ideas*, without revelation, to discover, whether Omnipotency has not given to some Systems of Matter fitly disposed, a power to perceive and think, or else joined and fixed to Matter so disposed, a thinking immaterial Substance: It being, in respect of our Notions, not much more remote from our Comprehension to conceive, that GOD can, if he pleases, superadd to Matter a Faculty of Thinking, than that he should superadd to it another Substance, with a Faculty of Thinking; since we know not wherein Thinking consists, nor to what sort of Substances the Almighty has been pleased to give that Power, which cannot be in any created Being, but merely by the good pleasure and Bounty of the Creator. For I see no contradiction in it, that the first eternal thinking Being should, if he pleased, give to certain Systems of created senseless matter, put together as he thinks fit, some degrees of sense, perception, and thought. . . . What certainty of Knowledge can any one have that some perceptions, such as . . . pleasure and pain, should not be in some bodies themselves, after a certain manner modified and moved, as well as that they should be in an immaterial Substance, upon the motion of the parts of Body: Body as far as we can conceive being able only to strike and affect body; and Motion, according to the utmost reach of our *Ideas*, being able to produce nothing but Motion, so that when we allow it to produce pleasure or pain, or the *Idea* of

a Colour, or Sound, we are fain to quit our Reason, go beyond our *Ideas*, and attribute it wholly to the good Pleasure of our Maker. For since we must allow he has annexed Effects to Motion, which we can no way conceive Motion able to produce, what reason have we to conclude, that he could not order them as well to be produced in a Subject we cannot conceive capable of them, as well as in a Subject we cannot conceive the motion of Matter can any way operate upon? (Locke, 1975, 4.3.6)

Locke argues that because of the nature of our concepts of mind and body we cannot rule out any hypothesis on the grounds of *a priori* reasons, even the hypothesis that some suitably organized material systems cannot be superadded with the power of thinking. But Locke also argues that dualism is not compellingly or independently justified. He also says that we cannot conceive how a separated mental substance could interact with the body, and he ends the paragraph showing that the formulation of the problem itself makes it entirely unintelligible:

He that considers how hardly Sensation is, in our Thoughts, reconcilable to extended Matter; or Existence to any thing that hath no Extension at all, will confess, that he is very far from certainly knowing what his Soul is. 'Tis a Point, which seems to me, to be put out of the reach of our Knowledge: And he who will give himself leave to consider freely, and look into the dark and intricate part of each Hypothesis, will scarce find his reason able to determine him fixedly for, or against the Soul's Materiality. Since on which side so ever he views it, either as an unextended Substance, or as a thinking extended Matter; the difficulty to conceive either, will, whilst either alone is in his Thoughts, still drive him to the contrary side. An unfair way which some Men take with themselves: who, because the unconceivableness of something they find in one, throw themselves violently into the contrary Hypothesis, though altogether as unintelligible to an unbiased Understanding. (Ibid.)

I do not want to further analyze Locke's text to establish whether we can abstract from the theological frame in which he argues the thinking matter hypothesis, or if the urgency to put the mind-body problem aside reflects a radical and unconvincing skepticism.

Beyond the *intentio auctoris*, for us to understand the history of the effects not only of the thinking matter hypothesis, but of Locke's picture of subjectivity as a whole, is crucial.

Many scholars are carrying out this work. On the one hand, John Yolton (1991) reconstructed the reception of the thinking matter hypothesis. He holds that Locke's hypothesis would eventually lead to the redefinition in dynamic terms of the old "inert" mechanistic concept of matter. On the other hand, a growing number of studies (Von Glasersfeld, 1992; Di Francesco, 1998; Holt,

1999) recognized the continuity between Locke's antisubstantialist turn and its contemporary version.

I tried to focus on some aspects of Locke's thought that have received limited attention in the critical literature, such as his desubstantialized view of mind and the critique of the monism-dualism dilemma. The lesson we could receive from Locke's philosophy of mind is that the desubstantialization of mind does not necessarily imply the return of the dilemma within it. Instead, we can begin to see an independent refutation of the monism-dualism dilemma as a presupposition of a more complex, rich, and articulated view of mind and of its relations with nature.

6. Diagnosis of the Reasons of a Dilemma: John Searle and Conceptual Dualism

Dualism and monism have a long philosophical history. Plato, Descartes, Popper, Thomas Hobbes, Quine, and Dennett are examples of authors who, in different times and with different modalities, made a definite choice within the scope of the traditional theoretical options (but see Zanet, 2007, for an interpretation of Quine that makes his thought continuous with the desubstantialization of mind). On the contrary, other philosophers began to see the monism-dualism dilemma as the root of a paradigm which, with its presuppositions and categories, prevents a richer and more articulated conception of the mind and of its embodiment in nature instead of contributing to its development.

I want to locate the theoretical proposal worked out by Searle with this naturalistic and non-reductive line of thought, with Locke's thought as its genetic moment in the postCartesian era.

Searle analyzed and criticized what he calls "conceptual dualism," a set of presuppositions on which the monism-dualism dilemma relies in contemporary philosophical debate. He states his critique in light of the most recent achievements of the philosophical and scientific debate, but although the compelling force and efficacy of the Chinese room argument is usually recognized (see for example, Crane, 2003, pp. 123–128; Preston and Bishop, 2002; Smith, 2003), the analysis of conceptual dualism and biological naturalism (his non-reductive model of the ontology of mind) did not receive the same interest. The reason is probably that philosophers saw the problem as already solved by functionalism, so that Searle's model would be at best a form of idiosyncratic redundancy with respect of the contemporary mainstream (Kim, 1995; Chalmers, 1998; Honderich, 2001).

But this interpretation does not take into account the situation of the contemporary debate. Monism and dualism, if my analysis is correct, returned within a desubstantialized conception of mind with their traditional problems, the problems of consciousness and mental causation.

According to Searle, the debate is structured by a set of categories and assumptions implicitly shared by all the participants (monists and dualists) but

quite probably false: the mutual ontological exclusion between physical and mental, the standard theory of causation, the ambiguity of the notion of reduction, and the ambiguity of the notion of identity (Searle, 2004, pp. 108–111, 115–125).

Taken together, the mentioned assumptions define conceptual dualism: It is customary to think of dualism as coming in two flavors, substance dualism and property dualism; but to these I want to add a third, which I will call “conceptual dualism.” This view consists in taking the dualistic concepts very seriously, that is, it consists in the view that in some important sense “physical” implies “nonmental” and “mental” implies “nonphysical.” Both traditional dualism and materialism presuppose conceptual dualism, so defined. I introduce this definition to make it clear why it seems to me best to think of materialism as really a form of dualism. It is that form of dualism that begins by accepting Cartesian categories. I believe that if you take those categories seriously . . . as a consistent dualist, you will eventually be forced to materialism. Materialism is thus in a sense the finest flower of dualism. (Searle, 1992, p. 26)

Monism and dualism would be logical consequences of a wider form of dualism the basic assumption of which is the mutual ontological exclusion of mental and physical. But what does “mutual ontological exclusion” mean?

Mental and physical are predicates of entities or properties defined in such a way as to make their conjunction self-contradictory. Materialism, Searle writes, states that we can bridge the Cartesian ontological gap through a reduction of the mental to the physical. But the history of materialism shows that this attempt failed. Materialism ends up asserting not that the mental *qua* mental (with its qualitative and subjective dimension) is part of the physical, but (more or less explicitly) that the mental *qua* mental does not exist; that only physical facts, described in a “third-person” terminology, exist (Searle, 2004, p. 127).

In this sense, materialism would be “the finest flower of dualism.” Far from dissolving the old Cartesian dichotomy of mental and physical, it accepts that the mental as such is problematic and that we have to deny features like subjectivity and qualitiveness to give a picture of mind that coheres with what we know about the rest of the natural world. But this stance would implicitly accept the Cartesian assumption that the mental is ontologically different from the physical.

So according to Searle, that the mind-body problem is still with us in its traditional form is not surprising, since the mutual exclusion principle is so widespread that it is also built into ordinary language:

[A]long with the Cartesian tradition we have inherited a vocabulary, and with the vocabulary a certain set of categories, within which we are historically conditioned to think about these problems. The vocabulary is not innocent, because implicit in the vocabulary are a surprising number of theoretical claims that are almost certainly false. The vocabulary includes a series of ap-

parent oppositions: “physical” versus “mental,” “body” versus “mind,” “materialism” versus “mentalism,” “matter” versus “spirit.” Implicit in this opposition is the thesis that the same phenomenon under the same aspects cannot literally satisfy both terms. Sometimes the semantics and even the morphology seem to make this opposition explicit, as in the apparent opposition between “materialism” and “immaterialism.” (Searle, 1992, p. 14)

I need to state in more detail the properties defining mental and physical as incompatible and the reasons supporting this opposition.

In the following chart (Searle, 2004, p. 116), we can see the system of oppositions to which I referred:

Mental	Physical
Subjective	Objective
Qualitative	Quantitative
Intentional	Non intentional
Nonspatially located; nonextended in space	Spatially located and extended
Not explainable by physical processes	Causally explainable by microphysics
Incapable of acting causally on the physical	Acts causally; causally closed system

For the sake of accuracy, I must point out that dualism and antireductionism describe the last two features of the mental (inexplicability by physics and causal incapability) in a weaker way. Chalmers, Levine, and Davidson think for example that the explanatory gap could be accidental because it would just be a feature of our descriptions that doesn’t affect the ontological structure of the events.

We could say the same about mental causation. Apart from Thomas H. Huxley, Jackson, and Chalmers, few philosophers explicitly espouse epiphenomenalism, although many dualistic conceptions imply it as a logical consequence. Leibniz’s pre-established harmony, for example, makes mind an epiphenomenon, since harmony itself does all the causal work. Thus, my arm goes up because God harmonized my will with my bodily movement, but in this view, no causal connection between intention and action exists.

Searle’s thesis is that some of the properties listed in the chart do not exist, while other oppositions are pseudo contradictions. For example, he writes that the distinction between quality and quantity is probably fictitious (ibid., p. 118). Clinical practice has made it possible to measure some qualitative conscious states. Thanks to the electroencephalogram, we can determine, for example, whether a subject is in REM phase sleep and consequently, whether a person is dreaming.

Analogously, some tests measure brain activity or measure times and modalities of reaction to stimuli, determining the quality of visual or auditory experiences for a subject. Also, some pointers based on the same standards can determine intensity and quality of an emotion (Damasio, 1994; 1999).

This does not mean that qualitative subjective mental states do not exist, but just that although a medical doctor cannot observe or feel in a first-person sense,

the pains of a patient, the proper theoretical and technological tools allow the scientific study and treatment of patients' subjective experiences.

We could say the same about the opposition between location in space and time for the physical and the phenomenological absence of these features for the mental (what should it mean that my belief that it is raining is three centimeters in length?). We could think that attributing a spatial extension to a mental event is, phenomenologically speaking, absurd, or that it is a categorical mistake.

We could say, with Walter Freeman (Searle and Freeman, 1998, p. 721) that a mental process is like a dancer's pirouette: trying to reduce a complex process to a single spatial place means making a misleading hypostatization.

Every process has a physical realization as its causal basis. This is true for pirouettes and mental states, as we can see from the studies made with the imaging techniques and from the studies on brain damage. But this does not mean that we can identify mental processes with their neural causal bases.

Consider the case of Phineas Gage (Damasio, 1994), a young worker who suffered an impressive work-related accident; a metal bar pierced his skull and brain at the level of the prefrontal lobes. He remained conscious and alive, but his personality changed dramatically. Formerly an honest young worker, he became a vicious and volatile person.

In this sense, even though the assertion that Gage's goodness and sociality resided in his prefrontal lobes sounds like an extreme oversimplification, we can say that these brain areas were responsible for the correct management of his rational control of emotions and behavior.

If this analysis is correct, we can perhaps begin to say that subjective experiences can be a legitimate subject matter of scientific study. The explanatory gap exists *today* because of contingent difficulties like the extreme empiricism described by Tononi. The case of Phineas Gage is again a good example: Even moral qualities such as sociality and being a good worker, like cognitive abilities and practical rationality, depend in large part on the correct functioning of some brain areas, although concepts like prefrontal lobe and morality (or emotion) may appear too different to be in any way related. What we do not know is why some brain areas are causally responsible instead of others and how brain electrochemistry produces the qualitateness and subjectivity of conscious life. But the concept of mind does not contain anything that we cannot scientifically investigate.

If Searle's analysis is correct, then the immediate conclusion would be that materialism is just as false as dualism because the reductionist structure defining it derives from the implicit assumption of a (paradoxically) unjustified Cartesian view of mind. Ultimately, we need to revise the conceptual and terminological apparatus that we use to state and discuss the mind-body problem.

A second issue defining conceptual dualism is the following: We cannot conceive of how the brain could be the causal basis of conscious mental states under the presuppositions of the mutual exclusion principle and the standard (nomological) theory of causation. According to this view, a causal relation holds only between events sequentially ordered in time and its exemplification instantiates

some regularity described by strict (universals) or *caeteris paribus* (constrained by initial conditions) laws. Working against the background of this conception of causality and of the exclusion principle, dualism and monism will encounter different difficulties.

Dualism describes consciousness as a non-physical event and it will have serious difficulties in giving an account of how we could have causal relations between separated ontological realms. Materialism, on the contrary, cannot characterize the relation as causal without implying the existence of mental events as distinct from brain events. The question is relevant because in a naturalistic context “being real” means first being in a network of causal relations.

Searle’s critique of the standard theory of causation is analogous to the argument against the principle of ontological exclusion between the mental and the physical. The actual practice of science requires us, in the case of the distinction between the mental and the physical, to work out a wider notion of the physical allowing for the inclusion of its qualitative component. In the same way, and for the same reason, an analysis of causation requires a wider notion of causality, not limited by the standard theory. Searle argues his thesis in passages like the following:

Lots of causal relations are between discrete events ordered in time. A paradigm case, much loved by philosophers, is the case of billiard ball one striking billiard ball two and stopping, while billiard ball two moves away. But not all causation is like that. In lots of cases of causation the cause is simultaneous with the effect. Look at the objects around you and notice that they are exerting pressure on the floor of the room you are in. What is the causal explanation of this pressure? It is caused by the force of gravity. But the force of gravity is not a separate even. . . . Furthermore, there are lots of cases of simultaneous causation that are, so to speak, bottom-up, in the sense that lower-level microphenomena cause higher-level macrofeatures. Again look at the objects around you. The table supports books. The fact that the table supports books is causally explained by the behavior of the molecules. (2004, pp. 123–124)

This thesis broadens the notion of causation and covers the cases in which cause and effect are not separate events but simultaneously realized.

The reason why explanations like “The table is solid because it has a given kind of molecular structure,” or “The book falls down because the force of gravity acts on it” are examples of causal explanations is that they exemplify cases in which *X* makes *Y* happen and, *a contrario*, the existence of *Y* is explained by *X*, so that we can say, counterfactually, that if *X* did not exist, then *Y* would have not happened.

On the contrary, the traditional model of causation would say that these are not cases of causal explanation, since gravity is not an event and no time gap exists between solidity and the molecular structure.

This analysis of causation is the core of biological naturalism, Searle's model of the ontology of mind, which non-reductively naturalizes mind and allows for a definite overcoming of the monism-dualism dilemma.

Two

BIOLOGICAL NATURALISM: A NATURALISTIC AND NON-REDUCTIVE ONTOLOGY OF CONSCIOUSNESS

In the first chapter, I surveyed the debate in which John R. Searle develops his philosophy of mind. The critique of functionalism and the more general critique of conceptual dualism play roles in the context of a debate based on the traditional monism-dualism dilemma. Notwithstanding the paradigmatic turn toward the desubstantialization of mind, the dilemma, together with the problems of consciousness and mental causation, returned even in functionalism.

Searle pointed out and criticized conceptual dualism, which is responsible for the renewed central role of the traditional problems of consciousness and mental causation. In addition, he worked out a model of the ontology of mind, called biological naturalism. We can interpret this theory as a version of the desubstantialization of mind in a new naturalistic framework provided by the critique of conceptual dualism and justified on the basis of the explanatory structures exhibited by scientific research.

The thesis that I argue is that the traditional problems of consciousness and mental causation do not emerge for biological naturalism by virtue of its critique of conceptual dualism.

Here I begin to argue and test one of the main hypotheses driving the present work. If my analysis is correct, we could reasonably assert that Searle's model is fruitful from the theoretical point of view because it can give efficacious solutions to the traditional philosophical problems. It also allows for a dialogue with the most recent scientific research. Searle's model is not only compatible with scientific research, but also, it gives a general model of the ontology and functioning of mind that can work as a general interpretive framework for scientific research.

In this chapter, my aim is the reconstruction, analysis, and justification of biological naturalism and the presentation of the aspects of consciousness that Searle thinks as essential to it: qualitateness, subjectivity, and unity with their internal relations.

1. Identifying the Target: the Structural Features of Consciousness

Many scholars think that trying to define the concept of consciousness is a step with many theoretical difficulties: words like "sensitivity," "sensation," and "awareness," usually used in common sense definitions, involve a pre-understanding of the concept so that these definitions will be unavoidably circular.

Consider the following definition: “Consciousness: The having of perceptions, thoughts, and feelings; awareness. The term is impossible to define except in terms that are unintelligible without a grasp of what consciousness means” (Sutherland, 1989).

According to Searle, the definition of consciousness would not be a serious difficulty, because we have to measure a definition within the context of the stage of research in which it is proffered and with the aims suitable to that stage. If we just want to start a research, then we need a pre-theoretical definition that we can revise in light of new developments and that allows us to pick out a family of phenomena (the *explananda*). Other kinds of pre-theoretical definitions have the same problems. If, for example, we define heat as the sensation produced in a subject when it hits a hot object, then we have a circular definition. While if we define water as a tasteless, colorless, and odorless liquid, then we have a vague definition (other substances have these features but are not water and also water can be in a solid or a gaseous state).

We usually solve these problems in another stage of our research, when the development of more detailed theories gives analytical definitions for the family of phenomena pre-theoretically defined, sometimes allowing us to revise the pre-theoretical definition. So, when on the basis of those theories, we redefine heat as medium kinetic energy of molecular motion, we eliminate the circularity problem. Likewise, when we redefine water as a chemical substance whose molecular structure is H_2O , we solve the vagueness problem. In light of these reflections, Searle gives a pre-theoretical definition of consciousness:

By “consciousness” I mean those states of sentience or awareness that typically begin when we wake up in the morning from a dreamless sleep and continue throughout the day until we fall asleep again. Other ways in which consciousness can cease is if we die, go into a coma, or otherwise become “unconscious.” (1998, p. 40–41)

As we can see from the following passage, this definition can cover a wide range of phenomena:

Think, for example, of the differences between these experiences—the smell of a rose, the taste of wine, a pain in the lower back, a sudden memory of a fall day ten years ago, reading a book, thinking about a philosophical problem, worrying about income taxes, waking up in the middle of the night filled with aimless anxiety, feeling a sudden rage at the bad driving of other drivers on the freeway, being overwhelmed by sexual lust, having pangs of hunger at the sight of exquisitely prepared food, wishing to be somewhere else, and feeling bored while waiting in a line. All of these are forms of consciousness, and though they were chosen precisely to illustrate variety, they do not begin to exhaust the actual varieties of conscious experiences. Indeed, during all our waking lives, as well as when dreaming

during sleep, we are in one or more forms of consciousness, and the conscious states have all the variety of those waking lives. (Ibid., p. 41)

Notwithstanding the variety of phenomena, we can identify some structural features of consciousness: qualitateness, subjectivity, and unity.

The problem of consciousness, according to Searle, is in part that of finding the relationships between those features and the rest of nature.

A. Qualitateness

Every conscious experience has a qualitative character, the “what is it like” that Thomas Nagel describes in his famous paper (1981): We feel “something like” tasting a beer, listening to a concert, thinking that $2+2=4$, seeing a rose, or being a bat. Each one of these experiences has a qualitative character distinguishing it from the others, but qualitateness as such is a general feature of all conscious states. This is the reason why, according to Searle (1998b; 2002b; 2002c), consciousness and qualitateness are coextensive. We cannot treat the problem of consciousness leaving out the so-called problem of qualia. Doing this would mean changing the subject matter.

In Ned Block’s terminology, we could say that Searle is arguing against the mutual independence of access consciousness and phenomenal consciousness. We think of the second, but not the first, as consisting of qualia. Here is Block’s definition of phenomenal consciousness:

P-consciousness is experience. P-conscious properties are experiential properties. P-conscious states are experiential states, that is, a state is P-conscious if it has experiential properties. The totality of the experiential properties of a state are the “what it is like” to have it. Moving from synonyms to examples, we have P-conscious states when we see, hear, smell, taste, and have pains. P-conscious properties include the experiential properties of sensations, feelings, and perceptions, but I would also include thoughts, wants, and emotions. (Block, 1997, p. 380)

Access consciousness, on the other hand, includes states with propositional content whose correct functioning in a cognitive organization is independent of the phenomenal realization of it:

A state is A-conscious if it is poised for direct control of thought and action. To add more detail, a representation is A-conscious if it is poised for free use in reasoning and for direct “rational” control of action and speech. (Ibid., p. 382)

In these definitions, *A* and *P* conscious states are not mutually exclusive categories: some *A* conscious states could be *P* conscious and a *P* conscious state could

be available for the rational control of the activity of an organism. So, perhaps, something like being deeply immersed in reasoning exists, while the sensation of a sudden noise could make me decide, rationally and deliberately, to interrupt my reflections to see if something dangerous is happening in my surroundings.

Block thinks also that these forms of consciousness are mutually independent: An organism could have *A* consciousness but not *P* consciousness.

Michael Tye is even more radical, since he holds that states with propositional content (intentional states) lack phenomenal aspects, which properly belong to sensations and emotive reactions:

Consider my desire to eat ice cream. Is there not something it is like to have this desire? If so, is not this state phenomenally conscious? And what about my belief that I am a very fine fellow? Or the memory that September 2 is the date on which I first fell in love? Is there not some phenomenal flavor to both of these states? In the former case, some phenomenal sense of pride and ego, and in the latter some feeling of nostalgia?

It seems to me not implausible to deal with these cases by arguing that insofar as there is any phenomenally or immediately experienced felt quality to the above states, this is due to their being accompanied by sensations or images or feelings that are the real bearers of the phenomenal character. Take away the feelings and experiences that happen to be associated with the above states in particular cases, and there is no phenomenal consciousness left. (Tye, 1995, p. 4)

Searle's critique goes against the idea that *A* consciousness is independent of *P* consciousness and that *A* conscious states do not have phenomenal aspects:

Some people think that qualia are characteristic of only perceptual experiences, such as seeing colors and having sensations such as pains, but that there is no qualitative character to thinking. As I understand these terms, that is wrong. Even conscious thinking has a qualitative feel to it. There is something it is like thinking that two plus two equals four. There is no way to describe it except by saying that it is the character of thinking consciously "two plus two equals four." But if you believe there is no qualitative character to thinking that, then try to think the same thought in a language you do not know well. . . . However, the point must be trivial; that is, whether or not conscious thoughts are qualia must follow from our definition of qualia. As I am using the term, thoughts definitely are qualia. (Searle, 2002c, p. 40)

Here Searle makes two points. First, thoughts (mental states with propositional content) are qualitative states, because something like thinking the same content in different languages with different levels of fluency in the two languages exists. Qualitativeness cannot be limited only to the sensory and emotive aspects of mind.

Second, the whole debate on qualia would rest on the way we define words such as qualia, phenomenal, and qualitativeness. Searle uses consciousness and qualia interchangeably because every conscious state has a qualitative aspect, so that making a distinction between consciousness and qualia would be misleading. Within the range of conscious experiences, we could instead make the distinction between intentional and non-intentional consciousness, both with qualitative features:

Conscious states always have a content. One can never just be conscious, rather when one is conscious, there must be an answer to the question, “What is one conscious of?” But the “of” of “conscious of” is not always the “of” of intentionality. If I am conscious of a knock on the door, my conscious state is intentional, because it makes reference to something beyond itself—the knock on the door. If I am conscious of a pain, the pain is not intentional, because it does not represent anything beyond itself. (Searle, 1992, p. 84)

Many authoritative voices take this thesis as controversial (cf. Tye, 1995; Crane, 2001b) since they think that we can analyze mind entirely as intentional representation of objects and states of affairs, including even apparently non-intentional phenomena like pains and undirected forms of anxiety and anguish. But even in Searle’s position, the distinction between intentional and non-intentional is less sharp than in the quoted passage, to which he adds this note:

There is one qualification to this point. The sense of body location does have intentionality, because it refers to a portion of the body. This aspect of pains is intentional, because it has conditions of satisfaction. In the case of a phantom limb, for example, one can be mistaken, and the possibility of a mistake is at least a good clue that the phenomenon is intentional. (Ibid., p. 251)

The debate among the supporters of the different positions is still open (cf. Crane 2001b, chap. 3). But Searle picks out a crucial element when he says that the discussion is based on different definitions of the concepts involved. So Tim Crane (2003b, p. 218) says that the meaning of the word “qualia” is ambiguous among different senses. First, having qualia is having experiences with a phenomenal character. Second, qualia are non-intentional features of experiences.

Crane (2003a, pp. 130–134) also holds that we could think that the debate depends on our more general theory of the mind-body relation, since reductionist and antireductionist philosophers use qualia terminology in different ways.

Daniel C. Dennett (1991, chap. 12; 2002), for example, attributes to antireductionism a conception of qualia that is identifiable with the Cartesian view that mind is acquainted with its products (ideas) only. On the basis of this interpretation, he refutes the antireductionist theses and eliminates the qualitative aspects of mental life. They would be by-products of mistaken philosophical views such

as the dogmas of incorrigibility and introspection: taking seriously the concept of qualitativeness on this interpretation means reviving a “Ghost in the Machine” (cf. Dennett, 2002).

At this point, we could (for example cf. Chalmers, 1996, p. 15) functionally define learning, identifying it with a set of brain processes producing stable changes in the behavior of the organism. The function of learning is to regulate the organism-environment adaptive relation. In this perspective, the sensations of effort and the conscious thoughts involved in learning processes remain unexplained or are eliminated. Why would some set of functionally defined brain processes give rise to qualitative phenomena?

But many philosophers, such as Searle and Crane, claim that no link between the concept of qualitativeness and the described Cartesian conception exists, so that qualia would not imply the dogma of the Ghost in the Machine. If their analysis is correct, the validity of non-reductive arguments is preserved and qualitativeness would be one of the aspects characterizing consciousness.

B. Subjective and Objective

Searle argues the thesis that subjectivity is an aspect of the nature of mind and consciousness through an analysis of the concepts of “subjective” and “objective.” This analysis has two aims. The first is to pick out the difference between the epistemic and the ontological sense of the distinction between subjective and objective. In this way, Searle points out a feature of consciousness that distinguishes it from the rest of the natural world.

Second, the analysis allows Searle to refute the thesis that the objectivity of science prevents scientific study of the subjective domain of mental states.

With respect to the first point, here is Searle’s analysis:

Epistemically speaking, “objective” and “subjective” are primarily predicates of judgments. We often speak of judgments as being “subjective” when we mean that their truth or falsity cannot be settled “objectively,” because the truth or falsity is not a simple matter of fact but depends on certain attitudes, feelings and points of view of the makers and the hearers of the judgment. . . . In the ontological sense, “objective” and “subjective” are predicates of entities and types of entities, and they ascribe modes of existence. In the ontological sense, pains are subjective entities, because their mode of existence depends on being felt by subjects. But mountains, for example, in contrast to pains, are ontologically objective because their mode of existence is independent of any perceiver or any mental state. (1995b, p. 8)

So, on one hand, the distinction has an epistemic sense. We apply the terms to judgments and the function of these words is to characterize different degrees of independence of truth values with respect to individual opinions and values. In

the ontological sense, we apply the terms as predicates of entities and they characterize their form of existence. In this sense, subjective and objective are categories of the empirical reality. The existence of entities like pains depends on their being experienced by a subject (an I, an agent), while the existence of other entities does not in any sense depend on the experience an agent or observer has of them, so their existence is objective or in the third person.

By virtue of this analysis, Searle argues that from a logical point of view, the argument that the objectivity of science prevents the study of the subjective consciousness suffers from an ambiguity of the terms involved. The reasoning does not follow because the subjectivity of consciousness is ontological, while the objectivity of science is epistemic. Also, we already have epistemically objective sciences of ontologically subjective phenomena, such as medicine and neuroscience:

The point of making this distinction is to call attention to the fact that the scientific requirement of epistemic objectivity does not preclude ontological subjectivity as a domain of investigation. There is no reason whatever why we cannot have an objective science of pain, even though pains only exist when they are felt by conscious agents. . . . Though many philosophers and neuroscientists are reluctant to think of subjectivity as a proper domain of scientific investigation, in actual practice we work on it all the time. Any neurology textbook will contain extensive treatment of the etiology and treatment of such ontologically subjective states as pains and anxieties. (Searle, 2002c, pp. 43–44)

C. Holistic Unity

The “unity of consciousness” is the subject of a complex debate in the history of philosophical and scientific thought (cf. Cleeremans 2003; Bayne and Chalmers, 2003; Tononi, 2003b; Di Francesco, 1998). René Descartes, for example, used the unity of consciousness (interpreted as indivisibility of the *cogito* in independent entities) to argue substance dualism: Since the *res cogitans* is indivisible and the *res extensa* is divisible, then mind and body are independent things.

David Hume, on the contrary, considered this unity a “grammatical fiction,” the creation of an apparently unified bundle of perceptions or ideas atomistically distinct and independent, where extrinsic relations of causality and resemblance create the appearance of unity and bind the ideas each other.

The concept of unity of consciousness worked out by Searle (1992, pp. 129–130; 2002b; 2002c, p. 41; 2004, chap. 5) differs from the Cartesian view and from Hume’s conception of the unity of the “I” as a fictitious bundle made of independent perceptions, while it refers to Immanuel Kant’s concept of “the transcendental unity of apperception” and to Gerald M. Edelman’s neurobiological theory of consciousness as a “remembered present” (cf. Searle, 1992, pp. 129–130).

According to Searle's view, each conscious state is part of a field, which is holistically unified according to two dimensions that he names "vertical" and "horizontal" for the synchronic and diachronic aspect of the unity of consciousness.

In its vertical dimension, consciousness unifies qualitatively different experiences at each given time. So, in this moment, I am reflecting on philosophical problems; I am writing; I am seeing the computer in front of me; I am feeling the pressure of clothes on my body; I am smelling some scents coming from the kitchen; I have a lot of other experiences unified in a single conscious field.

In its horizontal dimension, each conscious experience maintains its typical unity across time. According to Searle's example, I could not think a given thought, or issue its corresponding speech act, if iconic and short-term memory did not maintain its unity. This is a general feature of consciousness, since each conscious state is "internal" in a double sense:

Consciousness necessarily occurs inside an organism or some other system. Consciousness is also internal in a second sense, and that is that any one of our conscious states exists only as an element in a sequence of such states. One has conscious states such as pains and thoughts only as a part of living a conscious life, and each state has the identity it has only in relation to other such states. My thought, for example, about a ski race I ran long ago, is only that very thought because of its position in a complex network of other thoughts, experiences, and memories. My mental states are internally related to each other in the sense that in order for a mental state to be that state with that character it has to stand in certain relation to other states, just as the whole system of states has to be related to the real world. . . . Thus, the ontology . . . involves their being part of a sequence of complex conscious states that constitutes my conscious life. (Searle, 1998b, pp. 41–42)

Searle's model of unity is a holistic one: mental states do not come as independent ideas, but are interdependent. Each conscious experience, according to this model, is identifiable as such only in virtue of its position within a field of consciousness (1992, pp. 129–130). We cannot have isolated experiences that at another time are associated via a binding device: Each experience necessarily occurs within a pre-existing holistically structured conscious field. The phenomenon is salient in visual perception, as we can infer from the discovery, made by the Gestalt psychologists, of the figure-background structure of the perceptual field. But the holistically structured unity would be a general phenomenon, which is necessary to the existence of each kind of conscious experience.

Consider also that the mechanisms underlying vertical and horizontal unity could be the same. According to Edelman, the construction of the identity of mental states across time requires categorizations and recategorizations worked out by memory systems of the brain. These categorizations continuously compare and revise the patterns embodied in the past with the ongoing construction of conscious states in relation to the value that they have for the organism. This is

the reason why Edelman talks about value-category systems. The consequence would be that vertical and horizontal dimensions of unity are different aspects of the same phenomenon (the activity of categorization of the brain) since the diachronic component works even in the construction of the synchronic unity. The present is “remembered” because the unity of different conscious states in a holistically unified field is made possible by the fruitful interaction between past categorical patterns and ongoing categorizations (Edelman, 1992, pp. 119–120).

Finally, according to Searle, qualitateness, subjectivity, and unity, qua aspects articulating the nature of consciousness, are logically related. Qualitateness implies subjectivity, which in turn implies unity:

I used to think that these three features, qualitateness, subjectivity, and unity, could be described as distinct features of consciousness. It now seems to me that that is a mistake; they are all aspects of the same phenomenon. Consciousness is by its very essence qualitative, subjective, and unified. There is no way that a state could be qualitative, in the sense I have introduced, without it also being subjective in the sense I have explained. But there is no way that the state could be both qualitative and subjective, without having the kind of unity I have been describing. You can see this last point if you try to imagine your present state of consciousness broken into 17 independent bits. If this occurred, you would not have one conscious state with 17 parts; rather, there would be 17 consciousnesses. . . . It is absolutely essential to understand that consciousness is not divisible in the way that physical objects typically are; rather, consciousness always comes in discrete units of unified conscious fields. (Searle, 2004, pp. 136–137)

2. Biological Naturalism

Qualitateness, subjectivity, and unity are the features articulating the essence of conscious experiences. These features are also logically related; they distinguish consciousness from other natural phenomena. Here is the problem: What kind of relationships hold between these phenomena and the rest of the natural world?

We will recognize in this problem of consciousness a version of the more general mind-body problem, which also covers the problems of mental causation and intentionality. How could a physical system like the brain refer to objects or events different from it? How could a mental state act causally in producing behavior under the presupposition of the causal closure of the physical?

The following presentation illustrates the core theses of Searle’s position:

- (1) Conscious states, with their subjective, first-person ontology, are real phenomena in the real world. We cannot do an eliminative reduction of consciousness, showing that it is just an illusion. Nor can we reduce consciousness to its neurobiological basis, because such a third-person reduction would leave out the first-person ontology of consciousness.

- (2) Conscious states are entirely caused by lower-level neurobiological processes in the brain. Conscious states are thus *causally reducible* to neurobiological processes. They have absolutely no life of their own, independent of the neurobiology. Causally speaking, they are not something “over and above” neurobiological processes.
- (3) Conscious states are realized in the brain as features of the brain system, and thus exist at a level higher than that of neurons and synapses. Individual neurons are not conscious, but portions of the brain system composed of neurons are conscious.
- (4) Because conscious states are real features of the real world, they function causally. (Ibid., pp. 113–114)

The theses gave rise to many replies from the materialistic and dualistic sides of the debate. Colin McGinn (1999), for example, wrote that irreducibility of subjective ontology to objective ontology makes the critique against conceptual dualism useless, so that the old mind-body problem would return in disguise. We would still have to know how to reconcile the subjective and the objective instead of the mental and the physical (see also Nida Rümelin, 2002).

On the contrary, Jaegwon Kim (1995) argues that biological naturalism with its supposed reconciliation of naturalism and antireductionism not only uses a “layered” model of reality shared by dualists and materialists (Popper and Eccles, 1981; Hofstadter, 1984, pp. 229–364; Dennett, 1969) and is, therefore, redundant, but also that Searle cannot even solve the problem of mental causation in a way that could be coherent with the assumptions of naturalism and antireductionism (Dennett, 1993a; 2005).

Kim also points out that the causal interpretation of psychophysical relations is hardly understandable if not incoherent and that it is strongly idiosyncratic with respect to the standard theory of causation (Kim, 1995, p. 191; see also Thompson, 1980).

Others tried to classify biological naturalism within the traditional theories, emphasizing only some aspects of the model. Douglas Hofstadter (1980) and Dennett (1993a; Dennett and Searle, 1997), for example, argue that Searle’s model is an example of an almost religiously inspired dualism and criticized its refusal of functionalism and reductionism. Sir John Eccles (1980), on the contrary, criticizes its materialism (the causal link relating brain and mind) and regarded biological naturalism as a form of identity theory.

Finally, others said that the model is vague, arguing that biological naturalism does not state sufficient reasons to say that mind could be realized only in biological systems. This kind of critique is widespread. Its latest statement is in Hofstadter (2007, pp. 28–31), where the author refers to Searle’s position without quoting any passage or even citing any book. William Bechtel (1988) and Michele Di Francesco (2000; 2002, p. 164; 2004, p. 121) gave another example of this critique. Bechtel describes Searle’s position as a “biological reductionism.”

My thesis claims that we can answer the objections concerning the vagueness of the model and of its refutation of conceptual dualism pointing out that biological naturalism can give efficacious solutions to the problems of mental causation and consciousness. As a consequence, if my analysis is correct, the “layered” model of reality and the desubstantialization of mind need the refutation of conceptual dualism in order to produce a coherent and theoretically powerful ontology of mind.

Searle is not saying or implying that only brains can cause mind, consciousness, and intentionality. The embodied and biological nature of mind gives some constraints to the study of the mind, but this does not mean that it will be impossible to develop some engineering project of duplication of those causal powers allowing the brain to produce mental phenomena.

[W]hen I say that consciousness is a biological process, a lot of people think that I am into neuronal chauvinism, or have a fetish for carbon-based molecules, or some such nonsense. But I am not saying that we might not be able to produce an artificial device that was also conscious using some other means. My point is that until we know how the brain does it we are in a poor position to try to do it artificially (Searle and Freeman, 1998, p. 720).

3. Emergent Properties

The core thesis of biological naturalism says that mental states are caused by and realized in the brain with cause and realization at different levels of organization of the system. Searle typically introduces the thesis with analogies with other non-suspicious, higher-level features, justifies it through its generality, and finally argues for it with an analysis of the explanatory procedures of scientific practice. Let us examine these points.

The presentation of the main thesis of biological naturalism is the *incipit* of *The Rediscovery of the Mind*:

The famous mind-body problem, the source of so much controversy over the past two millennia, has a simple solution. This solution has been available to any educated person since serious work began on the brain nearly a century ago, and, in a sense, we all know it to be true. Here it is: Mental phenomena are caused by neurophysiological processes in the brain and are themselves features of the brain. To distinguish this view from the many others in the field, I call it “biological naturalism.” Mental events and processes are as much part of our biological natural history as digestion, mitosis, meiosis, or enzyme secretion. (Searle, 1992, p. 1)

One of the direct consequences is that we have to consider mind as a part of our natural history, like other biological processes. This is a *leit-motiv* of Searle’s thought, as we can see from other passages:

Basic to our world view is the idea that human beings and other higher animals are part of the biological order like any other organisms. Humans are continuous with the rest of nature. But if so, the biologically specific characteristics of these animals—such as their possession of a rich system of consciousness, as well as their greater intelligence, their capacity for language, their capacity for extremely fine perceptual discriminations, their capacity for rational thought, etc.—are biological phenomena like any other biological phenomena. . . . *Consciousness, in short, is a biological feature of human and certain animal brains. It is caused by neurobiological processes and is as much a part of the natural biological order as any other biological features such as photosynthesis, digestion, or mitosis.* (Ibid., pp. 89–90)

Searle expresses his radical dissent with respect to the passive subjectivity of the Cartesian tradition by building the mind into the biological order of the natural world. He also points out the link between these biological processes, which allow an organism to perpetuate its existence and the power to generate its point of view. This coheres with some of the most advanced theories in neurosciences (cf. Edelman, 1989; 1993; Edelman, Tononi, 2000; Damasio, 1994; 1999), which, unlike the abstract conception of mind put forward by computational theories, are trying to continue the Darwinian project of interpreting consciousness as an indispensable device managing the interactions between the organism and its environment.

From the following more detailed characterization of psychophysical relationship, we can reconstruct Searle's position in more detail:

To summarize my general position, then, on how brain research can proceed in answering the questions that bother us: the brain is an organ like any other; it is an organic machine. Consciousness is caused by lower-level neuronal processes in the brain and is itself a feature of the brain. Because it is a feature that emerges from certain neuronal activities, we can think of it as an "emergent property" of the brain. An emergent property of a system is one that is causally explained by the behavior of the elements of the system; but it is not a property of any individual elements and it cannot be explained simply as a summation of the properties of those elements. The liquidity of water is a good example: the behavior of the H₂O molecules explains liquidity but the individual molecules are not liquid. (Searle, 1997, pp. 17–18)

Consciousness, Searle says, *qua* mental (qualitative and subjective) is part of the biological order of the natural world. Its existence is causally explained by the interactions between the elements composing an organic system (the brain) and it is realized at a level of description (of the same system) higher than that of the basic elements. Consciousness is a "causally emergent property" of the brain in the same sense in which digestion is an emergent property of the digesting apparatus, or liquidity is an emergent property of water.

An emergent property is a feature of the entire system but not of its basic elements and it is causally explained not as the simple addition of the parts, but from the causal interactions at the level of the basic components of the system itself. No H₂O molecule is in a liquid state, unlike the system as a whole. In the same sense although no neuron can think, speak, or feel sensations, the system as a whole has these abilities.

Emergentism has a long philosophical tradition, which began with the British philosophers Samuel Alexander and Charles D. Broad (Alexander, 1920; Broad, 1925). Over time, it has been associated with dualistic positions (by Karl R. Popper, for example) and recently to positions compatible with reductionism (Kim, 1993; 1998; Crane, 2001a; 2001b). This is probably the reason why Searle distinguishes between two senses of the concept of emergence, the first being compatible with naturalism but not with reductionism and the second being compatible with dualism:

[C]onsciousness is a causally emergent property of systems. It is an emergent feature of certain systems of neurons in the same way that solidity and liquidity are emergent features of systems of molecules. The existence of consciousness can be explained by the causal interactions between elements of the brain at the micro level, but consciousness cannot itself be deduced or calculated from the sheer physical structure of the neurons without some additional account of the causal relations between them.

This conception of causal emergence, call it “emergent1,” has to be distinguished from a much more adventurous conception, call it “emergent2.” A feature *F* is emergent2 iff *F* is emergent1 and *F* has causal powers that cannot be explained by the causal interactions of *a*, *b*, *c* If consciousness were emergent2, then consciousness could cause things that could not be explained by the causal behavior of the neurons . . . on my view consciousness is emergent1, but not emergent2 (Searle, 1992, p. 112)

According to Searle, what holds for liquidity, solidity, digestion, and other higher-level properties or processes also holds for all our conscious states: pains, perceptions, emotions, beliefs, desires, and moods are entirely caused by the behavior of the basic elements of the brain and are also realized at the system level as features of the brain. So stated, Searle’s model also allows for an interpretation in terms of supervenience:

On the account I have been proposing, mental states are supervenient on neurophysiological states in the following respect: Type-identical neurophysiological causes would have type-identical mentalistic effects. . . . On this characterization of the supervenient relation, the supervenience of the mental on the physical is marked by the fact that physical states are causally sufficient, though not necessarily causally necessary, for the corre-

sponding mental states. That is just another way of saying that as far as the definition of supervenience is concerned, sameness of neurophysiology guarantees sameness of mentality; but sameness of mentality does not guarantee sameness of neurophysiology. (Ibid., pp. 124–125)

The notion of supervenience, introduced in philosophy of mind by Donald Davidson (1980b), formalizes the intuition that, given two systems, if they are type-identical with respect to their physical properties, then they will necessarily exemplify the same mental properties. Kim formalizes the basic intuition as follows:

Mental properties *supervene* on physical properties, in that necessarily, for any mental property *M*, if anything has *M* at time *t*, there exists a physical basis (or subvenient) property *P* such that it has *P* at *t*, and necessarily anything that has *P* at a time has *M* at that time. (Kim 1998, p. 9)

Philosophers of mind distinguish many kinds of supervenience principles. Chalmers (1996, chap. 2), for example, distinguishes between local, global, natural, and logical supervenience. But supervenience is not sufficient, according to Searle, to give a coherent account of psychophysical relation because it does not allow for a causal interpretation (as in Kim's case), or it does, but at the expense of being redundant:

The brain processes do not, at the level of neuron firings, constitute consciousness; rather, the neuron firings at the lower level cause the higher-level or system feature of consciousness. But if that is right, and everything we know about the brain suggests that it is right, then the concept of supervenience adds nothing to the concepts that we already have, such concepts as causation, including bottom-up causation, higher and lower levels of description, and higher-order features being realized in the system composed of the lower-level elements. Yes, consciousness is supervenient on brain processes, but you still have to tell us how it works. (Searle, 2004, p. 149)

According to Kim, supervenience does not allow for a coherent causal formulation, as he argues in the following passage:

in general, the relation between basic properties and supervenient properties is not happily construed as causal. For one thing, the instantiations of the related properties are wholly simultaneous, whereas causes are standardly thought to precede their effects; second, it is difficult, perhaps incoherent, to imagine a causal chain, with intermediate links, between the subvenient and the supervenient properties. . . . What intermediary events could causally connect a mental event with its subvenient physical basis? Would such intermediates themselves be mental or physical? (Kim, 1998, p. 44)

Kim argues the thesis on the basis of the standard theory of causation: We have a time gap between cause and effect and this means that we have a sequence of discrete events connecting the physical and the mental. Therefore, we are justified in asking whether this sequence would be made up of physical or mental events. Whatever the answer, the psychophysical relation does not allow for a coherent and informative answer, under a causal interpretation. If we conceive the events in the chain as physical, then we would still have to explain how a sequence of physical events could produce the mental. But the question still remains if we choose the second horn of the dilemma. These considerations lead into Searle's justification of his causal analysis of psychophysical relation. Here is a typical example:

Suppose an animal gets a shortage of water in its system. The shortage of water will cause "saline imbalances" in the system, because the ratio of salt to water is excessive in favor of salt. This triggers certain activities in the kidneys. The kidneys secrete rennin, and the rennin synthesizes a substance called angiotensin 2. This substance gets inside the hypothalamus and affects the rate of neuron firings. As far as we know the differential rates of neuron firings cause the animal to feel thirsty . . . that is the sort of explanation of how the existence of conscious feelings of thirst fits into our overall world-view. All forms of consciousness are caused by the behavior of neurons and are realized in the brain system, which is itself composed of neurons. What goes for thirst goes for all forms of our conscious life whatever, from wanting to throw up to wondering how to translate the poems of Stephane Mallarmé into colloquial English. All conscious states are caused by lower-level neuronal processes in the brain. (Searle, 2004, pp. 112–113)

This model exemplifies a general explanatory device that relies on the core theories of the scientific world view, so that the model is not an *ad hoc* hypothesis. Searle argues for this point with an analysis of the explanatory structure of atomic theory and evolutionary biology. Here is how the explanatory structure is articulated in Searle's fine-grained analysis:

Suppose I wish to explain why this pot of water is boiling. One explanation, a left-right macro-macro explanation, would be that I put the pot on the stove and turned on the heat under it. I call this explanation "left-right" because it cites an earlier event to explain a later event, and I call it "macro-macro" because both *explanans* and *explanandum* are at the macro level. Another explanation—bottom-up micro-macro—would be that the water is boiling because the kinetic energy transmitted by the oxidation of hydrocarbons to the H₂O molecules has caused them to move so rapidly that the internal pressure of the molecule movements equals the external air pressure, which pressure in turn is explained by the movement of the molecule of which the external air is composed. I call this explanation "bottom-up

micro-macro” because it explains the features and behavior of surface or macrophenomena in terms of lower-level microphenomena. . . . In evolutionary biology, there are characteristically two levels of explanations, a “functional” level where we explain the survival of species in terms of “inclusive fitness,” which depends on the phenotypical traits possessed by the members of the species, and a “causal” level where we explain the causal mechanisms by which the traits in question actually relate the organism to the environment. . . . If you put these two levels of explanation together, you get the following result: Because the phenotype, as produced by interaction of the genotype with the environment, has survival value relative to the environment, the genotype survives and reproduces. Such, in very brief form, are the mechanisms of natural selection. (1992, pp. 87–89)

So the explanatory device common to physics and biology shows the possibility of analyzing the structure and functioning of the same phenomenon or system at different levels of description of its organization. We have causal explanations with events sequentially ordered across time at the macroscopic and microstructural level, but we can also put the explanations together and give a causal explanation of the existence of the macrolevel in terms of microstructure as cause, and system macrofeatures as effect without time gaps. When we have to explain boiling water or a plant turning its leaves to the sun, we use a form of explanation that Searle names *bottom-up micro macro no time gap* where cause and effect are simultaneously realized and the effect (macro) is realized as a macrofeature of the system made out of that microstructure (micro) that in turn explains the existence and causal powers of higher-level or system features.

The explanatory structure of atomic physics and evolutionary biology would then show the existence of forms of causation that outstrip the boundaries of the standard, nomological theory of causation. Unlike typical forms of causation across time, these explanations make reference to a *bottom-up, no time gap* causation.

A consequence of this generality claim is the deflation of the traditional mind-body problem through the critique of that monism-dualism dilemma that limits its possible solutions. Mindful beings are natural, physical beings *qua* mindful. If we accept the generality claim put forward by biological naturalism, then, we can say that conceptual dualism is unjustified:

Consciousness is a higher-level or emergent property of the brain in the utterly harmless sense . . . in which solidity is a higher-level property of H₂O molecules when they are in a lattice structure (ice), and liquidity is a similarly higher-level emergent property of H₂O molecules when they are, roughly speaking, rolling around on each other (water). Consciousness is a mental, and therefore physical, property of the brain in the sense in which liquidity is a property of systems of molecules. If there is one thesis that I would like to get across in this discussion, it is simply this: The fact that a feature is mental does not imply that it is non physical; the fact that a fea-

ture is physical does not imply that it is not mental. Revising Descartes for the moment, we might say not only “I think, therefore I am,” but also *I am a thinking being, therefore I am a physical being*. (Ibid., p. 14–15)

Just as no one would think of system macrofeatures as something separate from the system itself, so we should not think that consciousness is independent of its neuronal bases. Dualism is ruled out: If we had a scientific theory connecting different levels of the same system, then we would not consider the emergence of new properties as something mysterious. We do not have a unified neurobiological theory, but this could just be a problem, which is not necessarily a mystery. The point that biological naturalism highlights is that we have no logical or metaphysical obstacles in principle for the construction of a theory because we have, within natural sciences, conceptual tools capable of accounting for the phenomenon without reducing nor eliminating it (and this also rules out monism).

4. Mental Causation: Consciousness and the Creative Structure of the Mind-World Relationship

My analysis of Searle’s theses has produced thus far the following theoretical proposal: Conscious mind is a natural phenomenon, the properties of which (qualitativeness, subjectivity, and unity) are biologically embodied system features of the brain, caused by lower-level interactions according to the mechanism of bottom-up, no time gap causation. It follows, therefore, that consciousness is a real phenomenon and that, like all real phenomena, it is involved in causal relations.

As we know from the zombie argument and from Kim’s causal exclusion dilemma, each description of mind in terms of irreducible higher-level features implies some problems concerning the intelligibility of mental causation.

Consider Searle’s version of the zombie argument. Suppose that each one of my neurons is replaced with a functionally equivalent silicon chip. What are the logical possibilities coming from this scenario?

One logical possibility, not to be excluded on *a priori* grounds alone, is surely this: you continue to have all of the sorts of thoughts, experiences, memories, etc., . . . that you had previously. . . . In this case, we are imagining that the silicon chips have the power not only to duplicate your input-output functions, but also to duplicate the mental phenomena, conscious and otherwise, that are normally responsible for your input-output functions. . . .

A second possibility, also not to be excluded on any *a priori* grounds, is this: as the silicon is progressively implanted on your dwindling brain, you find that the area of your conscious experience is shrinking, but that this

shows no effect on your external behavior. You find, to your total amazement, that you are indeed losing control of your external behavior. . . . We imagine that your conscious experience slowly shrinks to nothing, while your external observable behavior remains the same

Now consider a third variation. In this case, we imagine that the progressive implantation of the silicon chips produces no change in your mental life, but you are progressively more and more unable to put your thoughts, feelings, and intentions into action. In this case, we imagine that your thoughts, feelings, experiences, memories, etc., remain intact, but your observable external behavior slowly reduces to total paralysis. (Ibid., pp. 66–67)

The aim of this thought experiment is to pick out the logical possibilities available to characterize the causal relations among brain, mind, and behavior. In the first case, the substitution of neurons with chips duplicates input-output brain-behavior functions and the causal powers of the brain allowing it to cause consciousness. The second case foresees the duplication of input-output functions but not the causal powers of the brain to produce consciousness. Finally, the third case describes the duplication of consciousness, but not its causal power over behavior.

We can conclude, therefore, that on the basis of the second case, behavior and causal relations do not state a sufficient condition for the ontology of mental states. On the basis of the third case, we see that they do not even state necessary conditions:

Ontologically speaking, behavior, functional role, and causal relations are irrelevant to the existence of conscious mental phenomena. Epistemically, we do learn about other people's conscious mental states in part from their behavior. Causally, consciousness serves to mediate the causal relations between input stimuli and output behavior; and from an evolutionary point of view, the conscious mind functions causally to control behavior. But ontologically speaking, the phenomena in question can exist completely and have all of their essential properties independent of any behavioral output. (Ibid., p. 69)

This is just another statement of the subjective ontology of mind and of its irreducibility to a third-person ontology. But here we find something else. Behavior is irrelevant for the ontology of mind, although epistemically speaking, behavior can give us information about the presence of mental states in a system, while from a causal point of view, consciousness causally acts to mediate input-output relations. So, strictly speaking, the point of the thought experiment is ontological and it does not deny mental causation even from a logical point of view, provided that we carefully distinguish between ontol-

ogy (what exists), epistemology (how we know what exists), and causation (what it does).

Searle also argues that we have empirical proof of the causal role of consciousness, such as the reports about patients suffering epileptic attacks. Despite some patients continue their routine activities under an attack, they show diminished discriminative and ultimately creative power. This leads Searle to a hypothesis about the causal and evolutionary role of consciousness:

Complex behavior can be pre-programmed in the structure of the brain, at least as far as we know anything about how the brain works in such cases. Apparently, once started, the activity can run its course even in a *petit mal* seizure. But normal, human, conscious behavior has a degree of flexibility and creativity that is absent from the Penfield cases. . . . Consciousness adds powers of discrimination and flexibility even to memorized routine activities. . . .

The hypothesis I am suggesting then is that one of the evolutionary advantages conferred on us by consciousness is the much greater flexibility, sensitivity, and creativity we derive from being conscious. (Ibid., pp. 108–109)

Searle argues a crucial point here: The embodiment of mind as a part of the biological make-up of an organism requires that the discussion of mental causation takes into account what happens, from an evolutionary point of view, when an organism develops consciousness. Searle's idea is that having consciousness serves to structure the organism-environment interactions creatively:

[C]onsciousness serves to organize a certain set of relationships between the organism and both its environment and its own states. And . . . the form of organization could be described as "representation" . . . we can say that in conscious perception the organism has representations caused by states of affairs in the world, and in the case of intentional actions, the organism causes states of affairs in the world by way of its conscious representations. If this hypothesis is correct, we can make a general claim about the selectional advantage of consciousness: Consciousness gives us much greater powers of discrimination than unconscious mechanisms would have. (Ibid., p. 107)

These considerations, together with the holistic structure of mind and the bottom-up causation mechanism, give to Searle's analysis of psychophysical relation not only a remarkable originality but also theoretical power, since it can be effectively confronted with some crucial problems in contemporary debate.

The biological embodiment of mind, its holistic structure, and bottom-up causation allow us to work out a solid theory of mind that, while it can face the reductionist challenge posed by Kim's causal exclusion dilemma, puts forward a constellation of concepts that cohere with some of the most advanced scientific

theories on consciousness, like those worked out by Damasio, Edelman, and Giulio Tononi.

Before we make this comparison, I will have to analyze and critically assess Searle's theory in more detail and understand *in primis* whether, and to what extent, it can coherently combine naturalism with antireductionism. If Searle is right, the illusion of conceptual dualism is grounded on our definitions and not on patterns of facts, so that an empirical neurobiological theory can dissolve the illusion.

5. Reduction and the Practice of Definition

Searle's analysis of the ontology of mind in terms of emergent properties, different levels of description-organization, and bottom-up causation would appear *prima facie* to involve reductionism. We could argue that, after all, according to Searle, a mental state is nothing but a brain state. The same happens, on the other side, for other emergent properties. Solidity and liquidity reduce to the behavior of molecules, since once we have the complete causal story of biological processes such as photosynthesis and digestion, nothing is "left out."

Searle anyway rejects this interpretation, providing instead an ontological reading of classical antireductionist arguments such as Nagel's and Jackson's on the basis of his analysis of the subjective-objective distinction.

The aim of these arguments, Searle says, is not to state how we know the world or what we know about it, but "what real features exist in the world" (ibid., p. 117). He concludes, "No description of the third-person, objective, physiological facts would convey the subjective, first-person character of the pain, simply because the first-person features are different from the third-person features" (ibid.).

The standard answer to these arguments goes as follows (Churchland, 2002; Dennett, 1991, pp. 398–406; 2005, chap. 5): If Mary or the expert on bats does not know that the experience of yellow or the experience of the world mediated by a system of echolocation is identical to the physical or functional state *N*, we cannot draw the conclusion that the identity in question is not valid. We could say, with an argument in Nagel's or Jackson's style, that since *X* knows water under the description "tasteless and colorless liquid" but not under the description " H_2O ," then water is not H_2O . Then, Nagel's and Jackson's conclusions do not follow from their arguments.

According to Searle, this reply misunderstands the point of the arguments: Materialism does not state sufficient conditions for the existence of mental phenomena. The epistemic objections are correct but they are irrelevant for the arguments concerning the ontology of mind. Here is Searle's reply:

The problem in Mary's case is not just that she lacks *information* about some other phenomenon; rather, there is a certain type of *experience* that she has not yet had. And that experience, a first-person subjective phenomenon, cannot be identical with the third-person, objective neuronal and functional correlates. The point about the epistemology, the information, is

just a way of getting at the underlying ontological difference. . . . The logical form of the arguments is this: I stand in a relation to certain entities, my experiences of colors. And the bat stands in a relation to certain entities, its experiences of what it feels like to be a bat. A complete third-person description of the world leaves out these entities, therefore the description is incomplete. (2004, pp. 97–98)

This is a crucial point required for us to understand Searle's irreducibility thesis: Pains, for example, are irreducible to neurophysiological states because they have a subjective ontology. But if we express the thesis this way, Searle appears committed to some kind of property dualism (Hodgson, 1994; Stoutland, 1994; Chalmers, 1996, p. 132).

What are the results of a theoretical proposal that, like Searle's, primarily aims to justify a project of reintegration of consciousness into the natural world on the basis of the generality of the proposed explanatory model? The reply is that the irreducibility of consciousness and of its features has "no deep metaphysical consequences for the unity of our overall scientific world view" (Searle, 1992, p. 122) because:

[t]he contrast between the reducibility of heat, color, solidity, etc., on the one hand, and the irreducibility of conscious states, on the other hand, does not reflect any distinction in the structure of reality, but a distinction in our definitional practices. (Ibid., p. 123)

Here is the argument for Searle's thesis. First, according to Searle, the concept of reduction is ambiguous between ontological and causal reduction. Ontological reduction is defined as:

Objects of certain types can be shown to consist in nothing but objects of other types. For example, chairs are shown to be nothing but collections of molecules. This form is clearly important in the history of science. For example, material objects in general can be shown to be nothing but collections of molecules, genes can be shown to consist in nothing but DNA molecules. It seems to me this form of reduction is what the other forms are aiming at. (Ibid., p. 113)

Causal reduction is defined as:

the existence and *a fortiori* the causal powers of the reduced entity are shown to be entirely explainable in terms of the causal powers of the reducing phenomena. Thus, for example, some objects are solid and this has causal consequences: solid objects are impenetrable by other objects, they are resistant to pressure, etc. But these causal powers can be causally ex-

plained by the causal powers of vibratory movements of molecules in lattice structures. (Ibid., p. 114)

According to Searle, the standard reductive practices would go from causal reduction (discovery of the causal bases of the entity that we want reduce) to ontological reduction (identification of the reduced entity or property with its causal basis). He describes this pattern as the general principle underlying ontological reduction:

The general principle in such cases appears to be this: Once a property is seen to be *emergent*, we automatically get a causal reduction, and that leads to an ontological reduction, by redefinition if necessary. The general trend in ontological reductions that have a scientific basis is toward greater generality, objectivity, and redefinition in terms of underlying causation. (Ibid., p. 116)

Searle then shows that the pattern applies to secondary qualities as much as to primary qualities and then he argues as follows:

Thus, when the surface feature is a subjective appearance, we redefine the original notion in such a way as to exclude the appearance from its definition. For example, pretheoretically our notion of heat has something to do with perceived temperature: Other things being equal, hot is what feels hot to us, cold is what feels cold. . . . But when we have a theory of what causes these and other phenomena, we discover that it is molecular movements causing sensations of heat and cold. . . . We then *redefine* heat and color in terms of both the subjective experiences and the other surface phenomena. And in the redefinition we eliminate any reference to the subjective appearances and other surface effects of the underlying causes. "Real" heat is now defined in terms of the kinetic energy of the molecular movements, and the subjective feel of heat that we get when we touch a hot object is now treated as just a subjective appearance caused by heat, as an effect of heat. . . . The same pattern works for primary qualities: Solidity is defined in terms of the vibratory movements of molecules in lattice structures, and objective, observer-independent features, such as impenetrability by other objects, are now seen as surface effects of the underlying reality. Such redefinitions are achieved by way of carving off all of the surface features of the phenomenon, whether subjective or objective, and treating them as effects of the real thing. (Ibid., pp. 119–120)

The point of the passage from causal to ontological reduction is a redefinition of pretheoretical notions in terms of their causes. The causal bases of reduced entities are included in the new definitions, while surface features, subjective and objective, are eliminated and treated as effects produced by underlying causal structures. But none of these definitions eliminates the epistemic bases giving us the presence of the *explanandum*. Even when we have studied the physics of

color, heat, and solidity, we will continue to feel heat, see the color of fire, and feel the solidity of the table.

The redefinitions add nothing to what we discovered through causal reduction, but they are useful, since “To get a greater understanding and control of reality, we want to know how it works causally, and we want our concepts to fit nature at its causal joints” (ibid., p. 121). We can partly achieve this aim with the elimination of subjective and objective surface features from the new definition.

But this operation does not work for consciousness for a trivial reason which does not depend on reality, but on the definitional practices we have so far examined. Pains, for example, have subjective features and objective causes, like heat and color. Can we causally reduce them? Pains are caused by brain states, they are fully explainable in terms of patterns of neuron firings and so the answer is yes. Can we make an ontological reduction and say, for example, that pains are nothing but some kind of brain state? We would have to follow the standard pattern of ontological reduction and then exclude subjective features from the concept of “pain” ontologically reduced. The operation would not work because our interest in using the concept of pain is to name a subjective experience with qualitative features and so we cannot apply a definitional pattern whose goal is partly to distinguish appearance from reality within a conceptual scope in which they are identical:

Part of the point of the reduction in the case of heat was to distinguish between the subjective appearance on the one hand and the underlying physical reality on the other. . . . But we can’t make that sort of appearance-reality distinction for consciousness because consciousness consists in the appearances themselves. *Where appearance is concerned we cannot make the appearance-reality distinction because the appearance is the reality.*

[C]onsciousness is not reducible in the way that other phenomena are reducible, not because the pattern of facts in the real world involve anything special, but because the reduction of other phenomena depend in part on distinguishing between “objective physical reality,” on the one hand, and mere “subjective appearance,” on the other; and eliminating the appearance from the phenomena that have been reduced. But in the case of consciousness, its reality is the appearance; hence, the point of reduction would be lost if we tried to carve off the appearance and simply defined consciousness in terms of the underlying physical reality. In general, the pattern of our reductions rests on rejecting the subjective epistemic basis for the presence of a property as part of the ultimate constituent of that property. . . . Consciousness is an exception to this pattern for a trivial reason. The reason, to repeat, is that the reductions that leave out epistemic bases, the appearances, cannot work for the epistemic bases themselves. In such cases, the appearance is the reality. (Ibid., pp. 121–122)

The irreducibility of consciousness does not depend on patterns of facts. The concept of consciousness, unlike the concepts of heat and color, do not fit the pattern of ontological reduction. The irreducibility of consciousness does not imply the existence of any ontological gulf between the mental and the physical, or between the subjective and the objective. Like any emergent property, consciousness is causally explainable within the right theoretical framework that re-constructs the causal story of the phenomena systematically.

Consciousness also differs from the other emergent properties because its concept names ontologically subjective experiences. In this case, our concern is for appearance itself. We would lose the point of having the concept of consciousness, Searle says, if we exclude subjectivity from its definition, as ontological reduction requires.

Ontological reduction fails for the same reasons as it does for other phenomena (Searle, 2004, pp. 120–121). For example, money and performances of Beethoven's Ninth Symphony are, in principle, reducible to molecular structures or sound waves, but if a bank, economist, or musician were to redefine the respective concepts in these terms, they would lose the sense of the concepts themselves.

6. The Causal Exclusion Dilemma: Terms of the Problem

Let us recall the general terms of the problem of mental causation in the most recent debate, so that I can provide the context to analyze the debate between Kim and Searle and so that we can understand whether biological naturalism can provide a reliable solution to the causal exclusion dilemma.

Kim formalized the problem of mental causation as consisting of two presuppositions (causal exclusion and supervenience). The supervenience argument, or causal exclusion dilemma, derives from the conjunction of the presuppositions under a non-reductive interpretation of supervenience. We can state the first presupposition as follows:

[T]he presence of two causal stories, each claiming to offer a complete and full causal account of a given event, creates an unstable situation requiring us to find an account of how the two purported causes are related to each other. This is the problem of “causal/explanatory exclusion.” (Kim, 1998, p. 65)

If we have two causal explanations for the same event, each one claiming sufficiency and completeness, we need to establish whether they are in some way compatible (perhaps because they could be part of a single causal story) or if we have to discard one of them since we do not want overdetermination of the event.

What appears to make overdetermination an unviable option in the case of mental causation is the principle of causal closure, which requires that given a physical event with its ancestors and successors in the chain of causes and effects, we will still have only physical events. A violation of the principle would

imply the incompleteness in principle of physics and a return to Cartesian interactionism (Chalmers, 1996, pp. 156–160; Kim, 1998, p. 40).

The presupposition of mind-body supervenience states that mental properties supervene on physical properties if and only if two physically identical systems will exemplify the same mental properties. Once we accept the presupposition of the multiple realizability of mind, we see that the relation is asymmetric (different physical systems can realize the same mental state), but supervenience provides a framework of minimal physicalism on the basis of which mind depends on the physical: mental properties are asymmetrically co-variant with respect to physical properties.

Kim states his argument in terms of properties, namely in terms of those elements characterizing events, objects, and states mentioned as causally relevant. For example, the stone as such is not causally responsible for breaking the glass; the responsibility is with its solidity.

My formulation in terms of properties, events, and states aims at taking into account the ongoing debate between those philosophers saying that properties are separately mentionable with respect to events (such as Kim) and others, like Donald Davidson, who think that properties are parts of the events.

With these premises, we can develop the exclusion argument (for the complete statement see Kim, 1998, chap. 2).

- (1) Let M be the mental property, event or state that cause M^* .
- (2) Under the presupposition of mind-body supervenience, M^* has a physical basis responsible for its exemplification, P^* .
- (3) Then M^* is exemplified because it has a cause in M , as stipulated, or because, by virtue of supervenience, P^* is exemplified.

But, under the assumption of the general exclusion principle, this point creates a tension between these two explanations of M^* :

M^* occurs because its supervenient basis P^* occurs, and as long as P^* occurs, M^* must occur no matter what other events preceded this instance of M^* —in particular, regardless of whether or not an instance of M preceded it. This puts the claim of M to be a cause of M^* in jeopardy: P^* alone seems fully responsible for, and capable of accounting for, the occurrence of M^* . (Ibid., p. 42)

- (4) Under the premises of supervenience and exclusion, M can cause M^* only if it cause the exemplification of P^* .

We can support this point, as Kim says, with the general supervenience principle, which states that the supervenient property variation depends on the subvenient (or basic) property variation. The headache, for example, vanishes because the

medicine acts on brain chemistry. But, on the ground of the supervenience assumption, we have:

- (5) On the ground of the supervenience assumption, M is exemplified because P is exemplified.

This threatens P^* with causal overdetermination. We started with the hypothesis that $M \rightarrow M^*$ and, to make it compatible with supervenience and exclusion, we concluded that P^* has to be caused by M . But if M has its subvenient basis in P , then the problem of overdetermination that we had with M^* (caused by M and P^*) returns with P^* (caused by M and P) and, in addition, if we accept $M \rightarrow P^*$ we will have the failure of causal closure of $P \rightarrow P^*$.

Therefore, we have the following horns of a dilemma:

- (1) If $M \rightarrow M^*$ exemplifies a causal relation, then we have the failure of $P \rightarrow P^*$ causal closure, since the only way the relation could hold implies $M \rightarrow P^*$.
- (2) If we do not accept the failure of causal closure and we do not assert $M \rightarrow P^*$, then the $M \rightarrow M^*$ level is epiphenomenal. The regularity between M and M^* would be explained by the causal work made by P on P^* and by the supervenience of M on P and of M^* on P^* .

During the last decades, neodualism, antireductionism, and functionalism were under attack. The objection concerned the problem of mental causation. For example, Chalmers admits that epiphenomenalism is a hardly avoidable result for a position that, like his neodualism, has to reconcile property dualism with the causal closure principle.

Another supporter of epiphenomenalism is Frank Jackson. He thinks that the epiphenomenal character of qualia is not a reason to deny their existence (2002, p. 278). Meanwhile, he asserts that epiphenomenalism is compatible with an evolutionary view of biology since we can have causally ineffective and even evolutionarily dangerous features as by-products of evolutionary positive traits. For example, the thick skin covering a bear allows the animal to easily recover from cold temperature, but the thick skin makes the animal slower.

Analogously, Jackson thinks that qualia are nothing but causally ineffective by-products of brain processes (ibid., p. 277). But the compatibility of Darwinism and epiphenomenalism is controversial and, speaking in more general terms, epiphenomenalism is almost unanimously regarded as too high a price to assert the existence and irreducibility of mind.

We can put forward the following considerations: First, to doubt the causal action of mental states is to put the view of ourselves as cognitive agents in crisis. Kim wrote that this ordinary view necessarily involves mental causation to make sense of notions such as “knowledge,” “perception,” “action,” and “reasoning” (1998, p. 31).

Second, if epiphenomenalism is a threat even for functionalism then the consequences could be devastating for cognitive science, which is grounded on that framework:

[F]unctionalism might turn out to be a form of epiphenomenalism, and the received . . . view of cognitive science as an autonomous special science, which generates its own distinctive law-based causal explanations at a higher, formal/abstract cognitive level, faces an imminent collapse. There is irony in the fact that this is precisely the conception of cognitive sciences that has been inspired and promoted by functionalism—in particular, the view that properties studied by cognitive sciences are second-order properties that abstract from the nitty-gritty physical/biological details of the cognitive systems that realize them. (Kim, 1998, p. 52)

Also, Dennett writes (1993b, pp. 447–452) that epiphenomenalism necessarily implies a return to the old Cartesian, solipsistic view of mind: Epiphenomenal qualia could not, by definition, have any effect on the physical, so that their existence could not be communicable.

Since qualia would be epiphenomenal, a statement concerning their existence such as “I am in pain” or “I have epiphenomenal qualia” could equally be produced by a zombie (see also Chalmers 1996, chap. 5, on the paradox of the phenomenal judgment).

Even if we grant the existence of epiphenomenal qualia, we could have a systematic discrepancy between a subject’s “physical” assertions, its qualia, and its beliefs. The statement “I have qualia,” then, cannot be a proof of their existence for the speaker or for the hearer. So epiphenomenalism is not a reasonably supportable option or, *pace* Jackson, a safe way to cohere dualism and causal closure.

A different strategy to conjugate antireductionism and mental causation is to separate the issue from the more general debate on the mind-body problem, as Jerry Fodor, for example, does. His idea is that our ordinary explanatory practices imply the attribution of causally effective mental states to others and to ourselves and that our explanatory practices would be more reliable and non-negotiable than the metaphysical speculations about mind-body identity, supervenience, and causal closure.

Tyler Burge states the thesis as follows:

What interests me more is the very existence of the worries. I think that they are symptomatic of a mistaken set of philosophical priorities. Materialist metaphysics has been given more weight than it deserves. Reflection on explanatory practice has been given too little. The metaphysical grounds that support the worries are vastly less strong than the more ordinary grounds we already have for rejecting them. (Burge, 1993, p. 97)

The emerging idea is the following: We do not need to understand the mechanisms underlying psychophysical relation because ordinary practice can give good explanations of behavior. The practice of explanation does not need further foundation, because its validity does not depend on metaphysical assumptions about the relation between mind and body. Physical and mental explanations do not depend on each other, nor do they interfere with each other:

I think we have reasons, just from considering explanatory goals and practice . . . to think that mentalistic and physicalistic accounts of causal processes will not interfere with one another. Part of the point of referring to mental states lies in explaining intentional activity that involves (or is identical with) physical movement. A man's running to the store is explained by his believing that his child would suffer without the needed medicine and by his decision not to wait on a doctor. . . .

It would be perverse to think that such mental events must . . . fill some gap . . . in the chain of physiological events leading up to and including the movements of his muscles in running. It would be perverse to think that the mentalistic explanation excludes or interferes with non-intentional explanations of the physical movement. I think that these ideas seem perverse not because we know that the mental events are material. They seem perverse because we know that the two causal explanations are explaining the same physical effects as the outcome of two very different types of inquiry. . . . The upshot of this reasoning is that we have no ground for assuming that the failure of mental causes to interfere in the physical chain of events must be explained in terms of mental causes consisting in physical events. Interference would be surprising, given antecedent assumptions about mental and physical explanation. So non-interference is in no need of explanation in ontological terms. (Ibid., pp. 115–116)

According to Kim, the deflationist strategies do not overcome the problem. Physical and mental explanations belong to different fields of discourse and each one of these fields is legitimate in producing its explanations, but in the case of mental causation the fields produce different explanations of the same event and claim to state causally sufficient conditions.

In cases like these, the relativity of explanations does not help us to escape the problem. We have to figure out the relationships between different fields and different explanations. We have two reasons to do this. First, if the explanation is valid then the *explanans* must be responsible for the *explanandum* under the causal aspect:

[I]f *c* (or a description or representation of *c*) causally explains *e*, *c* must be a cause of *e*. If my desire for a drink of water causally explains my body's movement toward the kitchen, the desire must really be a cause of the bod-

ily movement. I take this to be an untendentious and uncontroversial point. (Kim, 1998, p. 64)

Second, the rivalry between explanations requires a systematic account of the relationships between the mentioned causes. Are these relationships partial causes of the same event or the same cause at different levels? If the first, how could we justify the inclusion of mental events in a causal chain that starts and ends with physical events without violating the causal closure? If the second, what is the relation between the levels?

We cannot escape the problem, according to Kim, through the isolation of mentalistic explanations with respect to the mind-body problem. Burge is right when he writes that the threat of epiphenomenalism is analogous to the skeptical worries about the existence of reality and of other minds. But where Burge thinks that the philosophical issues require a “therapy” that separates them from misleading metaphysical premises, Kim thinks that though (perhaps) nobody takes skepticism seriously, this does not prevent that such doctrine acts as a challenge to get a deeper understanding of the functioning of knowledge. So (perhaps) nobody takes epiphenomenalism seriously. But just as in the case of skepticism, we still need to remove epiphenomenalism through a systematic study of psychophysical relation (Burge, 1993, pp. 102–103; Kim, 1998, pp. 32, 61–64).

Perhaps Burge and Lynne Rudder-Baker, committed to the (supervenientist, but not to the identitist) assertion of the dependence of the mental on the physical, are also committed to the failure of causal closure. In any case, they need a theoretical account to argue for their answers and their positions (Paternoster, 2002, pp. 62–64).

A variation of the deflationary strategy, presented by Burge, Rudder-Baker, Fodor, and Robert Van Gulick, is the so-called generalization argument. Usually, the argument differs from the previous deflationary strategies since it explicitly assumes supervenience or a similar layered model and then tries to account for psychophysical relation within it. The argument states, in light of this general model, that Kim’s dilemma, if valid, would threaten with extinction all the causal powers of any higher-level phenomenon, since the dilemma should generalize.

Kim’s reply is that the problem of mental causation does not generalize because physical (including biological and chemical) properties are subject to functional reduction and this model could “save” mental causation: Just as “being a hearth” is a functionally analyzable property that can be reduced to many physical realizers, so “being a belief,” or a desire, or an intention, is also a functionally reducible property (Kim, 1998, pp. 97–103).

So the problem would not generalize if we assume that the general relation between different levels of natural reality is reductive and that mind is subject to this reduction.

Searle refutes both assumptions. Ontological reduction is just a useful logical tool to produce successful predictions and not a device that denies the differ-

ences between levels of reality. We cannot reduce the ontological subjectivity of mind for reasons intrinsic to the logical pattern of reduction.

Kim wrote that the subjective and qualitative dimension of mind is not subject to his reductive pattern, so that consciousness necessarily would be, in his view, epiphenomenal (*ibid.*, pp. 17, 101–103, 119). This means that even if we accept Kim's thesis, the generalization argument would still apply to consciousness.

Perhaps, then, we can think that a generalization argument suitably integrated with a general non-reductive explanatory device is a problem for the validity of Kim's dilemma. If the dilemma does not hold, then we can argue that the conjunction of antireductionism, mental causation, and causal closure is not contradictory.

7. Biological Naturalism and Causal Exclusion Dilemma in the Debate between Kim and Searle

In this section, I will examine whether Searle's model can give an efficacious answer to the causal exclusion dilemma and begin to test the possibility and fruitfulness of a naturalistic and non-reductive ontology of mind and consciousness.

To summarize the argument so far, a non-reductive view that stays in the naturalistic field and does not want to conceive of mental states as sheer *abstracta*, or give up and embrace epiphenomenalism, has to take up the task of showing how mental causation could exist under the presupposition of the causal closure of the physical world.

Kim's dilemma led to the following conclusion: Once we admit consciousness and its irreducibility:

[T]hen for each hypothetical mental cause, it will be possible to associate a physical cause. But then mental causes do not serve; once we assume the causal closure of physics, they reduce to physical causes or they are not real causes (my translation). (Di Francesco, 2002, p. 114)

We then must choose between epiphenomenalism and overdetermination unless we do not want to return, as Kim does, to identity theory and reductionism.

Searle's reply to the argument is noteworthy within the debate because while other theoretical positions worked out by philosophers such as Chalmers, Popper, Jackson, and others, accept the dilemma and work within it, Searle identified in conceptual dualism (especially in the exclusion principle of mental and physical) the implicit premise making the argument work and determining its structure.

According to Searle's analysis, even if we say that a physical system has many levels of description of its organization, we cannot derive from this premise that higher-level features do not have real causal powers, even though they can be explained by lower levels. Without conceptual dualism, the exclusion dilemma would have to hold for any system property. But how could we say that the causal role of the solidity of a hammer or of a piston is unintelligible?

The monism-dualism dilemma, which does not allow us to conceive of consciousness *as such* as a part of the natural world, is hidden in the formulation of the causal exclusion dilemma.

Here is a typical example of mental causation in terms of Searle's biological naturalism. At time t_1 , I have a strong headache (call this m_1) which causes, at time t_2 , my intention to take aspirin (call this m_2). m_1 is caused by patterns of neuron firings (call them p_1) and realized at a higher level of description, while m_2 is caused by other patterns of neuron firings (call them p_2) and realized at a higher level of description.

The first horn of Kim's dilemma states that if we accept the causal closure of the physical, then the causation across time from p_1 to p_2 is sufficient to cause m_2 , so that the macro level of mental causation ($m_1 \rightarrow m_2$) is epiphenomenal: the headache did not cause my intention to take aspirin because this intention has a sufficient physical cause that excludes the mental state as a cause.

In the second horn of the dilemma, if m_1 causes m_2 , then it must cause p_2 , which has a physical cause in p_1 , so in this case, we have the failure of causal closure at the lower level and causal overdetermination on p_2 . This would threaten the intelligibility of mental causation and its compatibility with naturalism.

The same happens with top-down or downward causation: For example, my intention to take aspirin (m_1) causes my bodily movement (call this action A). Even in this case, m_1 will be caused by p_1 , that causes p_2 , that in turn causes A. But then p_1 is sufficient for p_2 , and m_1 is an epiphenomenon, or p_2 is caused by m_1 and we have failure of causal closure and causal overdetermination.

If we accept the logic of the dilemma, then the only way out is a return to reductionism: Mind is causally efficacious because it is identical (in the strong sense of type identity theory) to the physical. In that case, we would have to deny antireductionism.

How does Searle answer the dilemma? First, he asks us to consider some non suspicious cases of causal explanation of a system at different levels of description. Consider the example of an ice block: Its solidity is caused by molecular behavior. The same holds for water in a liquid state. In these cases, we have bottom-up no time gap causation. Suppose that we heat the ice block to make it liquid: We have three causal explanations of the same system and the same phenomenon, each one true of one level and consistent with the others. We heated the ice (top level, left to right across time), an increase in energy caused an increase in molecule movements (bottom level, left to right across time), and the molecules are moving over each other in a random fashion (bottom-up, no time gap):

Does this imply overdetermination? Not at all. The same system is being described at different levels. Furthermore my heating the ice causes the change in molecule movements. Does this downward causation imply "failure of causal closure at the lower level?" Not at all. *Top-down causation always works across time, and it works because the top level is grounded in the lower levels.* In short, the same system admits of different

causal descriptions at different levels all of which are consistent and none of which implies either overdetermination or failure of causal closure. Nor, by the way, does anything here imply that heat, liquidity and solidity are epiphenomenal. Much of the point of levelism is to recognize causally real levels of description where, in some cases at least, the higher-levels are both caused by the behavior of elements at the lower levels and realized in the system composed of those elements. (Searle, 1995a, p. 218)

We can see the parallelism between these cases and the example of mental causation since, in these non suspicious cases of causal explanation, top-down causation does not violate the causal closure principle. Top-down causation works across time and it only works because the higher level is grounded on the lower level via bottom-up causation. In the example, we have no epiphenomenalism.

We are talking about a single physical system determined as such by bottom-up causation, where we describe the functioning of the system in many causally effective levels. This is the reason why we are not talking about two separated causal chains and why we have no overdetermination or epiphenomenalism.

Let us apply this story to mental causation. We will have the three mentioned levels of explanation and downward causation. Let us begin with the last case: Like with the ice block, we do not have failure of causal closure because downward causation works across time in virtue of the higher level grounded on the lower level via bottom-up causation.

Talking about the functioning of the system at the micro or macro level across time does not mean identifying different and separated sets of causes because the existence of bottom-up causation explains the higher level. Causally speaking, consciousness is not over and above its neurobiological basis. This means that, according to Searle's argument, the dilemma lacks its foundation.

In the interpretation of antireductionism, the mistake of the dilemma would be that Kim's reading implies conceptual dualism. Mental and physical properties are different, but not in the sense of the principle of ontological mutual exclusion.

We are talking about the difference between levels of the same system, all causally real although irreducible, with their causal powers explained by the microstructural level. We do not have two separate causal chains, but just one system moving as a whole, the inter-level relations of which explain the higher level causal efficacy.

The failure of conceptual dualism radically deflates the problem, as we can argue with a generalization of the exclusion dilemma for all the higher-level properties. It would be implausible to say that the ability of an organism to reproduce itself, or the property of a table to support objects, implies epiphenomenalism or failure of causal closure:

There are supposed to be two problems about mental causation: First, how can the mental, weightless and ethereal, ever affect the physical world? And second, if the mental did function causally would it not produce causal

overdetermination? The way to answer these questions is to abandon the assumptions that gave rise to them in the first place. The basic assumption was that the irreducibility of the mental implied that it was something over and above the physical and not a part of the physical world. Once we abandon this assumption, the answer to the two puzzles is first that the mental is simply a feature (at the system level) of the physical structure of the brain, and second, causally speaking, there are not two independent phenomena, the conscious effort and the unconscious neuron firings. There is just the brain system, which has one level of description where neuron firings are occurring and another level of description, the level of the system, where the system is conscious and indeed consciously trying to raise its arm. Once we abandon the traditional Cartesian categories of the mental and the physical, once we abandon the idea that there are two disconnected realms, then there really is no special problem about mental causation. (Searle, 2004, pp. 209–210)

In the end, we could say that the persistence of the problem, far from involving a return to reductionism, would instead be a sign of the difficulties of contemporary culture in getting rid of the heavy Cartesian inheritance.

Three

FUNCTIONALISTIC MODELS OF CONSCIOUSNESS: DENNETT, CHALMERS, AND THE DESUBSTANTIALIZATION OF MIND

One of the aims of John R. Searle's analysis of psychophysical relation is to pick out and criticize the implicit premise articulating the contemporary debate on consciousness according to the monism-dualism dilemma.

This premise is conceptual dualism with its principle of mutual ontological exclusion: mental and physical denote mutually exclusive ontological categories. But conceptual dualism lacks rational justification, since scientific research exhibits an explanatory structure accounting for the emergence of ontologically "new" levels. The novelty of the levels consists in their showing properties that do not exist in the lower levels, but we can explain these properties through a theoretical account that shows how lower-level interactions causally account for the existence and causal powers of them as system macrofeatures.

This explanatory structure (bottom-up no, time gap causation) can explain the existence of these properties within a naturalistic framework without involving any ontological gap and without any kind of reductionist strategy.

In this chapter, I will argue for one of the main hypotheses of this work: the conceptual problems of monism and dualism in relation to the problem of consciousness have their root in conceptual dualism, so we can dismiss them within a new naturalistic framework operating beyond the traditional principle of ontological exclusion.

I will argue for this point through the reconstruction and critical assessment of the theoretical positions worked out by Daniel C. Dennett and David J. Chalmers. I will especially show the dependence of their models and problems from conceptual dualism and the capability of biological naturalism to criticize them on detailed issues and general frameworks.

1. Dennett's Eliminative Naturalism: Multiple Drafts Model vs. the Myth of the Cartesian Theater

I have so far, in some sense, taken for granted the *explanandum*, presenting the problem of consciousness in terms of the necessity to account for the emergence in nature of a "new" ontological level whose properties are not reducible to the basic levels.

In Searle's terminology, the problem concerns the possibility to account for the power of some organisms to produce and support unified fields of qualitative subjectivity: being conscious for an organism means having qualitative states (perceptions, emotions, thoughts, sensations) such that we can say that the organism feels something that it is like having them. We can also say that these states have a subjective ontology: their existence depends on the organism's power to produce and support them. No subject, no consciousness.

Each conscious state occurs as part of a holistically structured unified field characterized by a high degree of integration and differentiation. Qualitativeness, subjectivity, and unity of consciousness are logically related. If conscious experience exists, then a subject must have it and the experience in question is articulated by the relationships holding within the holistic field.

The problem is to discover the relations between these properties and the rest of nature.

Dennett's position denies this formulation. He sees this picture of consciousness as nothing but a theoretical illusion. Consciousness would just be a product of the persistence of bad philosophical metaphors hanging together in the idea of an inner "Cartesian Theater" in which, magically, electrochemistry becomes experience "observed" by a "boss" (the subject) that coordinates the bodily activity.

Dennett's opinion is that the brain has no Cartesian Theater, but just many specialized agencies capable of organizing provisional "coalitions" whose aim is to manage the interaction between the organism and its environment. Consciousness is nothing but a virtual, von Neumannesque (serial) machine implemented by a parallel architecture. The idea that we have some kind of phenomenal consciousness may be easily refuted on the basis of empirical proofs (on the parallel architecture of the brain) and through philosophical arguments. For example, the idea that experience would consist of ineffable, intrinsic, and irreducible qualities appears, in light of Ludwig Wittgenstein's private language argument, wholly meaningless. Roughly speaking, if consciousness is private, then conscious states cannot be expressed in a public language and, therefore, they cannot play a role in the public and linguistically articulated linguistic game of science. Also, if we follow a verificationist interpretation of science (Dennett, 1991, pp. 390, 403, 460–461), we can say that what is private cannot be verified with scientific tools and theories, so that we cannot say that private conscious states exist.

According to Dennett, the nature of conscious phenomena, although closely related to its complex biological and functional basis of realization, is largely dependent on the concepts we use to describe it:

[C]onsciousness, like love and money, is a phenomenon that does indeed depend to a surprising extent on its associated concepts. Although, like love, it has an elaborate biological basis, like money, some of its most significant features are borne along on the culture, not simply inherent, somehow, in the physical structure of its instances. So if I am right, and if I succeed in overthrowing some of those concepts, I will threaten with extinction

whatever phenomena of consciousness depend on them. Are we about to enter the postconscious period of human conceptualization? Is this not something to fear? It is even conceivable? (Ibid., p. 24)

The thesis has a double function in Dennett's theoretical position. On the one hand, it links the analysis of consciousness with Dennett's theory of the intentional stance. According to this theory, intentional states are not real phenomena inside people but just the result of the stance that the others take towards the behavior of some complex systems. To explain this behavior, people find useful to postulate mental states such as beliefs, desires, and intentions, but *de facto* they are just using a normative vocabulary to describe brute, causal, objective processes that they do not know (Dennett, 1987).

On the other hand, the concept-dependent nature of consciousness foreshadows the following thesis: the falsification of a set of concepts used to describe conscious experience and the presentation of an alternative theory with corresponding new concepts involve the introduction of a new field of inquiry and the elimination of the old one with its supposed phenomena. This alternative perspective is the Multiple Drafts Model, as opposed to the Myth of Cartesian Theater.

The thesis concerning the concept-dependent nature of mind is part of a wider naturalistic philosophical stance, described by Dennett (1995) as a form of *mild realism* derived from the connection between Gilbert Ryle's critique of Cartesian dualistic metaphysics and Willard van Orman Quine's ontological relativity thesis: to say that mind is some kind of "inner" and private "thing" (the Ghost in the Machine) is just a categorical mistake; instead the study of mind cannot be independent of the ontological and epistemological commitment of current science, with its objectivity and its third-person point of view.

We can reconstruct the Multiple Drafts Model starting with Dennett's self-interpretation:

There is no single, definitive "stream of consciousness," because there is no Central Headquarters, no Cartesian Theater where "it all comes together" for the perusal of a Central Meaner. Instead of such a single stream (however wide), there are multiple channels in which specialist circuits try, in parallel pandemoniums, to do their various things, creating Multiple Drafts as they go. Most of these fragmentary drafts are of "narrative" play short-lived roles in the modulation of current activity but some get promoted to further functional roles, in swift succession, by the activity of a virtual machine in the brain. The seriality of this machine (its von Neumannesque character) is not a hardwired design feature, but rather the result of a succession of coalitions of these specialists.

The basic specialists are part of our animal heritage. . . . The result is not bedlam only because the trends that are imposed on all this activity are themselves products of design. Some of this design is innate, and is shared

with other animals. But it is augmented, and sometimes even overwhelmed in importance, by microhabits of thought that are developed in the individual, partly idiosyncratic results of self-explorations and partly the pre-designed gifts of culture. Thousands of memes, mostly borne by language, but also by wordless “images” and other data structures, take up residence in an individual brain, shaping its tendencies and thereby turning it into a mind. (1991, p. 253–254)

Consider the core concepts of Dennett’s position as expressed in the last quotation. First, the fictitious nature of what Gerald M. Edelman (2004, pp. 5–8) defines as the most remarkable Jamesian property of consciousness: the unity of the stream of consciousness.

Dennett’s thesis says that the brain has no Cartesian Theater, no single place in which all information comes together. Nowadays, this thesis is widespread among cognitive scientists and supported by empirical proofs from local cognitive dysfunctions such as blindsight, prosopagnosia, and split brain. Research and clinical reports suggest that the brain has many specialized organs or agencies.

But Dennett goes further: If we have no Cartesian Theater then the unity of consciousness does not exist. Likewise the whole image of consciousness as a real, subjective, and qualitative phenomenon is false.

Second, these cognitive agencies produce a “parallel pandemonium” of data and information. This metaphor means that the specialized cognitive agencies are not hierarchically and serially well ordered. Instead, they anarchically fight for what Dennett called “cerebral fame,” or “fame in the brain” (1993c, pp. 929–931), where this means that a brain event becomes conscious when it assumes saliency with respect to the background information and it can play further functional roles in the modulation of the organism-environment interaction. These functional roles are typically the use of these contents to perform actions and, in the case of human beings, to give a linguistic report of them.

Third, the mechanism determining the “promotion” of a content is a von Neumannesque (serial) “virtual machine” whose realization is not hardwired (determined once and for all). It instead results from the composition of the provisional coalition of cognitive agencies that assumes the control of organism-environment interaction. As Edelman and Giulio Tononi say, we should think of our brains not as hierarchically organized structures, but as organs whose well ordered functioning results from chaotic dynamics, like a free playing orchestra that can find its order without a director:

[I]magine a peculiar (and even weird) string quartet, in which each player responds by improvisation to ideas and cues of his or her own, as well as to all kinds of sensory cues in the environment. Since there is no score, each player would provide his or her own characteristic tunes, but initially these various tunes would not be coordinated with those of the other players. Now imagine that the bodies of the players are connected to each other by

myriad fine threads so that their actions and movements are rapidly conveyed back and forth through signals of changing thread tensions that act simultaneously to time each player's action. Signals that instantaneously connect the four players would lead to a correlation of their sounds; thus, new, more cohesive, and more integrated sounds would emerge out of the otherwise independent efforts of each player. . . . Although no conductor would instruct or coordinate the group and each player would still maintain his or her style and role, the players' overall productions would tend to be more integrated and coordinated, and such integration would lead to a kind of mutually coherent music that each one acting alone could not produce. (2000, p. 49)

Fourth, the dynamics are not arbitrary. Two items are responsible for the regulation of the competition among contents, agencies, and coalitions: the "design" and the "memes."

The embedding of an agency within a coalition can be favored by the biological functions that our natural design, stored in the DNA, states for the agency.

A content can have a greater "echo" and "become famous" in a brain (playing further roles) because having it can lead to evolutionary advantages. This is the function of memes (units of cultural transmission of ideas, customs, values), that make a content more easily embeddable on the basis of its ability to survive in a given culture.

Consciousness is *defined*, in Dennett's theory, by the mechanisms stated in this conceptual framework, as we can see in this passage:

I hereby declare that YES, my theory is a theory of consciousness. Anyone or anything that has such a virtual machine as its control system is conscious in the fullest sense, and is conscious *because* it has such a virtual machine. (Dennett, 1991, p. 281)

Dennett thinks that the subject of experience is nothing but the virtual and provisional product of the winning coalition of agencies in competition. He writes:

By yoking these independently evolved specialist organs together in common cause, and thereby giving their union vastly enhanced powers, this virtual machine, this software of the brain, performs a sort of internal political miracle: it creates a virtual captain of the crew, without elevating any one of them to long-term dictatorial power. Who's in charge? First one coalition and then another, shifting in ways that are not chaotic thanks to good meta-habits that tend to entrain coherent, purposeful sequences rather than an interminable helter-skelter power grab. (Ibid., p. 228)

Functional and computational processes realized on multiple parallel channels do not produce, constitute, or form the basis of a unified phenomenal con-

sciousness. Concepts like qualitativeness and subjective character of conscious experience do not designate a class of phenomena for which we have to account. Instead they are a multifaceted illusion produced by bad theories (Dennett, 1993b; 1993c, p. 891).

Those bad theories presuppose the existence of phenomena falling under the concept of qualitative (or phenomenal) mental state and then try to reduce them to third-person, intersubjectively observable phenomena. Or, on the contrary, they try to argue for the irreducibility of subjectivity.

Dennett's theoretical strategy, which he inscribes in the eliminative tradition (2002, p. 244), consists of two arguments. First is the dependence of the concepts of qualitativeness and subjectivity on the Myth of the Cartesian Theater.

Second is the confused and self-contradictory nature of the concept of qualitativeness as defined by second-order properties (privacy, incorrigibility, and direct access) whose theoretical legitimacy is philosophically refuted by arguments such as Wittgenstein's on private language.

The idea grounding the metaphor of the Cartesian Theater is that although concepts such as "soul" and *res cogitans* are no longer viable options, the corresponding fundamental intuitions were reintroduced by some philosophical views that saw consciousness as a special sort of privileged *medium*. The medium would be a single place in the brain where the information comes together and gives life to the experiences in front of the Subject.

But the studies on parallel signal processing in the brain and about the absence of a single place for reception, processing, and distribution of "messages" in the brain refute the Myth of the Cartesian Theater, a Myth that also gives rise to pseudo-problems (such as the homunculus fallacy) that are not present in the Multiple Drafts Model.

Even if we grant the existence of some kind of Cartesian pineal gland, we still would have the problem of explaining the "magic transduction" from the objective electrochemistry to the subjective experiences.

The Myth is also implicit, according to Dennett, in the formulation of problems and theories made by many neuroscientists, cognitive scientists, and philosophers of mind.

We could say, for example, that Ray Jackendoff was left with the unnecessary mind-mind problem (what is the relationship between computational and phenomenal mind?) because he took seriously the concept of qualitativeness.

Another victim of the Myth could be Jerry Fodor's modular and computational model, since Fodor postulates a non-modular center in which all the information processed by peripheral modules comes together. The center would function as a coordinator of all higher-order mental activities:

According to Fodor [the modules] do not carry out whole tasks in the economy of the mind . . . but stop abruptly at an internal edge, a line in the mind over which they cannot step. There is a central arena of rational "belief

fixation,” Fodor claims, into which the modules slavishly deposit their goods, turning them over to *nonmodular* . . . processes. . . .

Moreover, Fodor claims (with curious satisfaction) that no branch of cognitive science . . . has any clue about how this central facility does its work! (Dennett, 1991, p. 260)

So the protagonists of the debate do not explicitly support the existence of the Cartesian Theater only because its presuppositions are already implicit in the formulation of problems and theories. Usually, instead, “Cartesian materialists” (as Dennett calls these scholars) resolutely refute the Myth because they are not aware that the denial itself is motivated by a widespread superficiality instead of a close examination of contemporary scientific achievements (Dennett, 1993c, p. 920).

In the end, Dennett’s message is that no *medium* exists. Consciousness is not a special transduction medium and we have no Subject, no “me” in the brain. If we say that this subject exists and that any objective scientific theory leaves it out, then we ignore a crucial feature of successful scientific explanations. Successful scientific explanations always leave something out:

If you propose a theory of the knowing Subject that describes whatever it describes as like the workings of a vacant automated factory—not a Subject in sight—you will seem to many observers to have changed the subject or missed the point. On the other hand, if your theory still has tasks for a Subject to perform, still has a need for the Subject as Witness, then although you can be falsely comforted by the sense that there is still somebody at home in the brain, you have actually postponed the task of explaining what needs explaining (Dennett, 2005, p. 144).

This argument introduces the second line of attack, the aim of which is to justify the elimination of qualia through a series of arguments, metaphors, and tales, the logical structure and results of which I can only synthesize in their logical structures and results here.

We can define qualia as follows:

“Qualia” is an unfamiliar term for something that could not be more familiar for each of us: *the ways things seem to us* . . . look at a glass of milk at sunset; *the way it looks to you*—the particular, personal, subjective visual quality of the glass of milk is the *quale* of your visual experience at the moment (Dennett, 2002, p. 226).

The concept is apparently non-suspicious, but Dennett’s general thesis is that we do not have good reasons to hold that these qualitative phenomena exist. If we analyze, for example, the presuppositions articulating the main arguments put forward by the supporters of qualia, we can see that the concept of quale is de-

finer by second-order properties whose existence is nowadays strongly doubted, if not fully discredited.

In Dennett's reading, qualia would just be another name for what seventeenth century philosophers named "ideas:" mental particulars available only via introspection, knowable only by acquaintance (and then ineffable), and incorrigible (a subject could never be wrong about the qualities of its experience). This conception would necessarily produce problems. As Richard Rorty (1982) notices, to postulate the existence of qualia is to put the clock back to the old solipsistic, intellectualized view of mind as consisting of "appearances" that prevent us a direct access to reality and give rise to the skeptical doubts about the existence of the external world and of other minds. This means, among other things, to ignore what the linguistic-pragmatic turn produced: as we noticed about Wittgenstein, the linguistic game of scientific knowledge, like all languages, has to be public because no private language exists. There is no room for private entities in the public space of language and knowledge.

Rorty (1982; 1984) sees Thomas Nagel and Searle as the main authors of the return to this Cartesian view of the mind, against the achievements of the pragmatic and linguistic turn of contemporary philosophy. I will deepen the issue in the next chapter, but here I note that Rorty does not properly consider the role played by the concept of "Background," that could be Searle's solution against a possible return to the Cartesian solipsistic subjectivism (Searle, 1979c; 1980; 1983, chap. 5).

Francesca Di Lorenzo Ajello (2001, pp. 57–66) argued for this point through a close examination of the theoretical path leading Searle from *Speech Acts* to *Intentionality*. She writes:

[Searle] insisted on the role that rules and conventions play in the determination of meaning and in the crucial role . . . played by the common and intersubjective Background of practices and capacities in which necessarily . . . every meaning and every mental states have their roots (my translation). (Di Lorenzo Ajello, 2001, p. 59)

2. The Reasons of the Phenomenal: Searle vs. Dennett

Based on my analysis of Dennett's position, I conclude that the desubstantialization of mind is compatible with eliminativism. Also, since the concepts of qualitativeness and subjectivity would be embedded in discredited theories of mind, we could say that eliminativism is directly implied by the desubstantialization: once we eliminate the *res cogitans*, we have no room for the existence of the conscious mind. We just have physical, objective processes responsible for the interactions with the environment.

The echo of Dennett's theses among cognitive scientists working on robotics and post-classic cognitive science (Clark, 1997) and the wide philosophical discussions of his work (Brook and Ross, 2002; Dalbhom, 1993; Elton, 2003,

Jackson, 1993; Rosenthal, 1993, Shoemaker, 1993) notwithstanding, many critics pointed out some crucial problems. Consider the first assertion of Dennett's theory: the stream of consciousness does not exist because we have no Cartesian Theater in the brain. But this argument is not valid (Block, 1993, pp. 187–188). The integration of multiple states in a single scene and the construction of a unified stream in which these scenes follow each other does not require that a single *medium* processes and realizes the scene itself. The thesis that the distributed processing implies the virtual nature of the stream of consciousness must still be argued.

Reports about dissociative syndromes (blindsight, hemispatial neglect, identity dissociation, alien thoughts in schizophrenia) and the thousands of unconscious processes that psychoanalysis and cognitive science consider as the basis of conscious mental life in some sense cracked, if not dissolved, the Cartesian picture of a substantial and fully self-transparent subjectivity.

But identifying realism about consciousness with the Cartesian view is probably excessive if not tendentious. As Searle and others said (Searle, 1992, pp. 144–149; 1997, p. 124; Di Francesco, 1998, pp. 53–64, 84ff.), in everyday life, we experience emotive conflicts, *akrasia*, and other phenomena whose conflicting nature does not destroy the unity or the reality of the subject. We could say the same about impairments such as blindsight and other cognitive syndromes. Alien hand syndrome, split brain, and alien thoughts are more depressing situations, but nothing prevents us from treating them as extreme, pathological forms of conflict that we ordinarily experience, like the weakness of the will: dissociative syndromes, *a contrario*, shed light on the reality and fundamental unity of consciousness instead of destroying it (Di Francesco, 1998; Liotti, 2000).

Michele Di Francesco expressed this point as follows:

Suppose that Dennett . . . succeeded in showing that there is no brain place where all information we call conscious comes together and that instead, all brain activity is highly parallel and distributed. The lack of a center at the level of material structures does not tell us anything about what happens at the phenomenological level—unless we *already* decided to identify consciousness and brain closely and that the *ontological region* in which we have to look for the subjects is that described by the language of neurobiology. In the same spirit, the fact . . . that there is no single place of consciousness as the object of *scientific* (objective, third-personal, located in space and time) inquiry does not mean at all that the unity of . . . experience that every one of us normally experiences has to be classified under the voice *illusion*. It is still possible to defend the idea that it is perfectly real *even though it is subjective*. The idea that reality means objectivity . . . is not an undisputable dogma (my translation). (1998, p. 55)

These considerations lead us to examine one of the core theses of Dennett's theory, the supposed dependence of the existence of phenomenal states on the Myth of Cartesian Theater.

Dennett replies to the attacks against his supposedly tendentious stance (Tye, 1993; Shoemaker, 1993) by asserting without further argument that only a lack of imagination prevents us from seeing the nexus between the concept of phenomenal consciousness and a theory of mind unanimously recognized as untenable.

But Dennett's argument presupposes the validity of a reductionist approach (the stream of consciousness is real if and only if it is identifiable with a single brain place) whose eliminative conclusions should not be presupposed by the argument. Dennett's argument, if this analysis is correct, is begging the question. This is not surprising because one of Dennett's aims is to shift the burden of proof on the opponents (2002, p. 227): the supporters of qualia have to explain why we should worry about qualia and take them seriously.

Many critics pointed out the tendentious stance of the argument stated to achieve the shift of the burden of proof. So, for example, Jaegwon Kim writes:

One problem with many of these arguments is that they try to force qualia to bear the burden of all the exaggerated claims ever made about their special features, with the unsurprising result that nothing *can* qualify as qualia. To believe in qualia, it is not necessary, for example, to insist on absolute first-person infallibility or third-person inaccessibility. (2006, p. 235)

Dennett briefly considers, but rejects, the possible critique in Kim's style. For example, he recognizes that no one today explicitly holds the dogma of incorrigibility (except perhaps Saul A. Kripke), but he also adds without argument that believing in qualia implies a superficial belief in incorrigibility (2003, p. 233, 245).

We have other reasons to think that Dennett's critique of the notions of qualitativeness and subjectivity does not have a reference. Tim Crane and Searle's analysis of the debate on qualia is largely terminological: the discussion would be partly dependent on the more general dispute on the mind-body problem. The contenders often present their theses with reference to the notion of quale most capable of supporting their arguments. This is one of the reasons that led Searle to get rid of the qualia jargon as redundant with respect to the concept of consciousness, which he defines on the basis of the features of qualitativeness, subjectivity, and holistic unity.

Di Francesco writes:

[W]e could (re)interpret Dennett's proposal as centered on two points: (1) there are no qualia in the sense of intrinsic qualitative states; (2) there is no problem of locating conscious experiences in the naturalistic world view. Dennett seems to hold that it is enough to argue (1) in order to defend also (2). However, it seems clear that they are independent theses whose relation cannot be taken for granted (my translation). (2002, p. 217)

Dennett's theses on the elimination of consciousness are subject to this analysis, as we can see from Searle's arguments on these issues.

Searle disputes the validity of the epistemic argument that consciousness is not a valid topic for scientific research on the basis of his analysis of the epistemic and ontological senses of the subjective-objective distinction. He also criticizes the idea that the concept of qualitative state depends on the notions of introspection, incorrigibility, and direct access.

Searle agrees with Rorty and Dennett about the genesis of the problem: it comes from Descartes' project to ground metaphysics and knowledge on the epistemic certainty about the *cogito*. But Searle's conclusions differ from those of Rorty and Dennett. He writes:

I believe that at least two . . . mistakes have a common origin in Cartesianism. Philosophers in the Cartesian tradition in epistemology wanted consciousness to provide a foundation for all knowledge. But for consciousness to give us a certain foundation for knowledge, we must first have certain knowledge of conscious states; hence the doctrine of incorrigibility. To know consciousness with certainty, we must know it by means of some special faculty that gives us access to it; hence the doctrine of introspection. . . .

In any case, several recent attacks on consciousness . . . are based on the mistaken assumption that if we can show that there is something wrong with the doctrine of incorrigibility or introspection, then we have shown that there is something wrong with consciousness. But nothing could be further from the truth. Incorrigibility and introspection have nothing to do with the essential features of consciousness. They are simply elements of mistaken philosophical theories about it. (Searle, 1992, p. 149)

The Rorty-Dennett line of thought wants to overcome Descartes saying things such as "mind is a blur" (Rorty, 1982) and, "if the body had been easier to understand, nobody would have thought that we had a mind" (Rorty, 1979, p. 239). Rorty (1970) means that the concept of mind is ill-formed because it mixes a heterogeneous set of phenomena and processes (sensations, judgments, reasoning, and moral responsibility) that have nothing in common but the false feature of first-person incorrigibility.

Searle argues that these conclusions rely on a conceptual confusion between subjective ontology and epistemic certainty (1992, p. 145). For example, we cannot be wrong on our *beliefs* to have a given mental state, but it does not follow from this point that we cannot be wrong about the content or the kind of psychological state presenting the content. We could be victims of self-deception, or we could misunderstand our mental states, even though we sincerely feel pain to a ghost leg, or we sincerely believe to love a person. We cannot be wrong about the feeling of pain, or about having a given belief (that we love that person), although the pain can be produced by a deviant causal chain and the belief can be false.

This means that the thesis that incorrigibility is an essential feature of phenomenal consciousness does not take into account that the model of error grounded on the distinction between appearance and reality is not the only one available. In some cases we can be wrong because something *appears* to be such and such while it is *really* otherwise; but misinterpretation of a mental state, for example, relies on the hermeneutic criterion of whole-parts agreement:

Just as a person may misinterpret a text by failing to see how the elements of the text relate to each other, and by failing to understand the operation of the Background circumstances in which the text was composed, so a person may misinterpret his own intentional states by failing to see their interrelationships and by failing to locate them correctly relative to the Background of nonrepresentational mental capacities. In such cases we do not have the traditional epistemic model of making incorrect *inferences* on the basis of insufficient *evidence*. It is not a question of getting from appearance to reality, but rather of locating a piece in a puzzle relative to a whole lot of other pieces. (Ibid., p. 148)

Regarding direct access and introspection, these supposed mental properties would be, according to Searle, nothing but bad philosophical metaphors worked out to understand subjectivity. The ordinary notion of introspection does not give rise to problems, since it reduces to human beings' power to examine and assess mental states. But the metaphor of looking inside oneself through a special and incorrigible mind's eye is completely misleading as a description of the activity in which agents capable of reflection are ordinarily engaged. Here Searle explains why:

In the case of vision, we have a clear distinction between the object seen and the visual experience that the perceiver has when he perceives the object. But we can't make that distinction for the act of introspection of one's own mental states. (Ibid., p. 144)

The metaphor of direct access fails for analogous reasons:

[F]or there to be something to which I have privileged access, I would have to be different from the space in which I enter. But just as the metaphor of introspection broke down when the only thing to be observed was the observing itself, so the metaphor of a private inner space breaks down when we understand that there isn't anything like a space into which I can enter, because I cannot make the necessary distinctions between the three elements of myself, the act of entering, and the space into which I am supposed to enter. (Ibid., p. 98)

In the end, we can say that once the mentioned conceptual confusions are eliminated, we have no reason to doubt the existence of qualitative and ontologically subjective mental states, since subjectivity and qualitateness are independent of the Cartesian model of epistemic certainty.

3. Chalmers's Neodualism

If Dennett's theory of consciousness shows the compatibility of the desubstantialization of mind and eliminative conclusions, Chalmers's philosophy of mind shows that the thesis can be consistent even with the second horn of our dilemma, namely dualism.

Chalmers thinks that we have to take consciousness and its subjective character seriously: working out a theory of consciousness means accounting for the subjective character of experience within a naturalistic framework. We cannot leave out subjectivity without changing the subject matter.

Reductionism, Chalmers says, was the preferred theoretical strategy pursued by cognitive models of consciousness, mostly concerned with functional processes responsible for cognitive abilities, most of which fall under the label of Ned Block's "access consciousness" or, in Chalmers's terminology, "cognitive consciousness."

None of these models usually concerned with cognitive abilities such as reflection and language can explain why those processes have a subjective dimension. Chalmers (1996, pp. 115–117) also makes analogous considerations concerning the neurobiological theories worked out by Francis Crick, Christof Koch, and Edelman. He says that these theories do not explain anything: they would merely state correlations between the neurobiological and the phenomenal, without any explanation of why the correlations hold.

The failure of cognitive and neurobiological models is, according to Chalmers, a clue that materialism cannot explain consciousness and that we must substitute it with a dualistic ontology.

The solution to the problem of consciousness would be a "naturalistic property dualism," that Chalmers justifies through an extension of the zombie argument. The basic intuition is that if all physical facts are not sufficient to imply the existence of consciousness (in Chalmers's terminology, if consciousness does not logically supervene on the physical), then the facts relative to the phenomenal are "further" facts.

Since we can conceive of a world physically identical to ours but with a zombie twin for each conscious individual in our world, we can argue for the conclusion that the failure of logical supervenience on the physical falsifies materialism and justifies dualism.

We can summarize Chalmers's theses as follows: functionalism and cognitive science can give reliable explanations of cognitive phenomena such as reflection, learning, and language; but they cannot explain conscious phenomena

because phenomenal consciousness does not logically supervene on the physical. Therefore, phenomenal consciousness is irreducible.

We can still think that consciousness naturally supervenes on the physical, and if this is right, we could reconstruct the contingent psychophysical bridge laws determining this supervenience.

To understand and properly assess Chalmers's theses, we have to further examine his theoretical presuppositions and fundamental concepts.

A. Cognitive and Phenomenal Consciousness; Easy and Hard Problems

Chalmers reinterprets Block's distinction between access and phenomenal consciousness as the distinction between "psychological" and "phenomenal" properties of consciousness. Functionalism and cognitive sciences can analyze psychological properties, but the existence of phenomenal properties remains a mystery. He illustrates the distinction in these terms:

The first [aspect of the mind] is the *phenomenal* concept of mind. This is the concept of mind as conscious experience, and of a mental state as a consciously experienced mental state. This is the most perplexing aspect of the mind and the aspect on which I will concentrate, but it does not exhaust the mental. The second is the psychological concept of mind. This is the concept of mind as the causal or explanatory basis for behavior. A state is mental in this sense if it plays the right sort of causal role in the production of behavior, or at least plays an appropriate role in the explanation of behavior. . . . On the phenomenal concept, mind is characterized by the way it *feels*; on the psychological concept, mind is characterized by what it *does*. (Ibid., p. 11)

Chalmers next describes different problems deriving from these two concepts of mind that had a wide echo among philosophers and cognitive scientists (Shear, 1995; Dennett, 2005). Psychological phenomena are a source of difficulty for cognitive science, but we can think that they are just technical problems and that we will eventually solve these issues with a functionalist account. The problems with psychological properties would be "easy," but phenomenal properties would give rise to the "hard" problem:

The phenomenal aspects of mind are a different matter. Here, the mind-body problem is as baffling as it ever was. The impressive progress of the physical and cognitive sciences has not shed significant light on the question of how and why cognitive functioning is accompanied by conscious experience. . . . The hardest part of the mind-body problem is the question: how could a physical system give rise to conscious experience? (Chalmers, 1996, p. 25)

To discuss and solve the hard problem, we first need a theoretical framework that fulfils the adequacy conditions stated by Chalmers (taking seriously consciousness, science, and naturalism). Supervenience can give a framework which is “minimal” enough to be compatible with many approaches, but it is also naturalistic enough to rule out spiritualistic and antiscientific views.

B. Logical Supervenience, Materialism, and Reductionism

Supervenience names an asymmetric relation of covariance between properties. We can analyze it within either an antireductionist (Donald Davidson) or reductionist (Kim) interpretation. Chalmers introduces the notions of logical and natural supervenience, identifying the first as the logical device underlying reductive explanations of the kind that we can find in the natural sciences (in Chalmers’s sense of reductive):

B-properties supervene *logically* on A-properties if no two *logically* possible situations are identical with respect to their A-properties but distinct with respect to their B-properties. . . . It is useful to think of a logically possible world as a world that it would have been in God’s power (hypothetically!) to create, had he so chosen. God could not have created a world with male vixens, but he could have created a world with flying telephones. In determining whether it is logically possible that some statement is true, the constraints are largely *conceptual*. The notion of a male vixen is contradictory, so a male vixen is logically impossible; the notion of a flying telephone is conceptually coherent, if a little out of the ordinary, so a flying telephone is logically possible. (Ibid., p. 35)

Logical supervenience of *B* properties on *A* properties is a relation constrained only by conceptual criteria defined in Kripke’s terminology of the logically possible worlds. Logical supervenience between two kinds of property holds if and only if we cannot coherently conceive of a possible world in which *A* properties, but not *B* properties, hold.

To use Chalmers’s example, biological facts logically supervene on physical facts, because once the first are fixed, biological properties will follow. Two physically identical worlds will also be biologically identical, as a matter of conceptual necessity:

At the global level, biological properties supervene logically on the physical properties. Even God could not have created a world that was physically identical to ours but biologically distinct. There is simply no logical space for the biological facts to independently vary. . . . If there is a living kangaroo in this world, then any world that is physically identical to this world will contain a physically identical kangaroo, and that kangaroo will automatically be alive. (Ibid.)

The notion of logical supervenience accounts for the conceptual link between materialism and reductive explanation because we can define materialism as the thesis that the universe consists entirely of physical facts. This means:

[M]aterialism is true if all positive facts about the world are globally logically supervenient on the physical facts . . . if materialism is true, then once God fixed the physical facts about the world, all the facts were fixed. (Ibid., p. 41)

If we add to this definition of materialism the thesis that reductive explanation requires logical supervenience, we have a conceptual nexus among logical supervenience, materialism, and reductionism, allowing for a reformulation of the hard problem of consciousness. Consider the redefinition of reductive explanation in terms of supervenience:

If the property of exemplifying a phenomenon fails to supervene logically on some lower-level properties, then given any lower-level account of those properties, there will always be a further unanswered question: Why is this lower-level process accompanied by the phenomenon? Reductive explanation requires some kind of analysis of the phenomenon in question, where low-level facts imply the realization of the analysis. So reductive explanation requires a logical supervenience relation. For example, it is precisely because reproduction is logically supervenient on lower-level facts that it is reductively explainable in terms of those facts. (Ibid., p. 48)

As Chalmers writes, we can always ask further questions (why are these microlevel phenomena accompanied by these macrolevel phenomena?) only when a conceptual link between *A* and *B* properties describable in terms of logical supervenience is not available.

Let us return to the problem of consciousness in the light of the notions of logical supervenience, materialism, and reductive explanation. Materialism necessarily implies the reduction of consciousness to physical facts (because we have no further facts) and reduction of consciousness in turn implies logical supervenience of consciousness on the physical.

If we refute the possibility to reduce consciousness, materialism fails: this refutation is the aim of Chalmers's version of the zombie argument.

C. Zombie Argument and Naturalistic Property Dualism

To assess Chalmers's version of the zombie argument, we need to start with his formulation of the theoretical context. The basic argument goes as follows:

- (1) In our world, there are conscious experiences.

- (2) There is a logically possible world physically identical to ours, in which the positive facts about consciousness in our world do not hold.
- (3) Therefore, facts about consciousness are further facts about our world, over and above the physical facts
- (4) So materialism is false (ibid., p. 123).

The first premise is not disputable because its denial involves a violation of Chalmers's first adequacy condition. The premises three and four follow from the second. Therefore, Chalmers has to argue for the second premise. He uses two strategies: an appeal to intuition to shift the burden of proof and the extension to the global level (sets of worlds) of thought experiments such as Block's "Chinese nation," in which the absence of qualia is limited to a single individual. The following passages illustrate both these strategies:

I confess that the logical possibility of zombies seems equally obvious to me. A zombie is just something physically identical to me, but which has no conscious experience—all is dark inside. While this is probably empirically impossible, it certainly seems coherent; I can discern no contradiction in the description. In some ways an assertion of this logical possibility comes down to a brute intuition. . . . Almost everybody, it seems to me, is capable of conceiving of this possibility . . . in general, a certain burden of proof lies on those who claim that a given description is logically *impossible*. (Ibid., p. 96)

[W]e can indirectly support the claim that zombies are logically possible by considering nonstandard realizations of my functional organization. My functional organization . . . can in principle be realized in all sorts of strange ways. To use a common example . . . the people of a large nation such as China might organize themselves so that they realize a causal organization isomorphic to that of my brain, with every person simulating the behaviour of a single neuron, and with radio links corresponding to synapses. The population might control an empty shell of a robot body, equipped with sensory transducers and motor effectors. . . . All that matters here is that the idea such a system lacks conscious experience is *coherent*. A meaningful possibility is being expressed, and it is an open question whether consciousness arises or not. (Ibid., p. 97)

The failure of logical supervenience of consciousness on the physical implies the failure of materialism, but Chalmers does not argue only for negative theses. Instead, he develops a view of consciousness called "naturalistic property dualism" based on the hypothesis that mind naturally supervenes on the physical: in our world, mind supervenes according to contingent laws. Here is Chalmers's description of his position:

We can use Kripke's image here. When God created the world, after ensuring that the physical facts held, *he had more work to do*. He had to ensure that the facts about consciousness held. The possibility of zombie worlds or inverted worlds shows that he had a choice. The world might have lacked experience, or it might have contained different experiences, even if all the physical facts had been the same. To ensure that the facts about consciousness are as they are, further features had to be included in the world. (Ibid., p. 124)

He also writes:

The dualism implied here . . . is a kind of property dualism: conscious experiences involve properties of an individual that are not entailed by the physical properties of that individual, although they may depend lawfully on those properties. Consciousness is a feature of the world over and above the physical features of the world. This is not to say it is a separate "substance"; . . . All we know is that there are properties of individuals in this world—the phenomenal properties—that are ontologically independent of physical properties. (Ibid., p. 125)

Phenomenal consciousness is ontologically independent of the physical, but realized in our physical world through nomological connections: phenomenal properties are different in type with respect to physical properties.

Scientific research can try to discover nomological connections between phenomenal experience and its physical substrate to describe *how* experience arises in our world. But if we accept the validity of the zombie argument, then even with a complete physics and neurobiology, we will never overcome the conceptual gap between physical and phenomenal properties. This point brings us back to property dualism: the reason *why* experience arises in our world remains a mystery.

4. The Reasons for the Unity of the Natural World: Searle vs. Chalmers

Sandro Nannini provides an excellent exemplification of the main critique of Chalmers's approach:

It appears disputable . . . that from the *logical* possibility of zombies one could derive such a strong *ontological* conclusion as that consciousness is irreducible to the physical world. Such conclusion, in fact, derives from that premise only under the condition that one has a Platonistic and absolute view of the possible worlds, according to which they are in God's mind before creation and are not changeable. It seems really more plausible to consider them as theoretical constructs, the plausibility of which depends on the conceptual system, historically changeable, that we adopt. The existence of zombies appears possible today because the actual concept of con-

sciousness (and mind) embedded in ordinary language is a Christian and Cartesian one at its roots . . . (2002, pp. 155–156)

Chalmers's argument that the conceivability of a zombie world implies the failure of materialism is not compelling because of two closely related reasons. First, Chalmers draws ontological conclusions from epistemic premises. Second, he does not take into serious consideration that historical-cultural experiences, including scientific theories, constrain logical conceivability.

For example, if we believe in the existence of an explanatory gap between microphysics and macrolevel phenomena, our belief does not imply that light is not a matter of electromagnetic reflections or that water is not identical with a system of hydrogen and oxygen molecules.

We could say that Chalmers is returning to a version of the old Cartesian argument to justify the ontological independence of mind and body. René Descartes held that substance dualism follows from the possibility of conceiving of the soul surviving the body. Chalmers, on the other hand, thinks that his naturalistic property dualism follows from the logical possibility of zombie worlds: we could coherently subtract conscious experience from the world while its physical features remain the same.

Chalmers (1996, pp. 132–133) denies the identity between his position and Descartes' argument because the zombie argument concerns supervenience and not identity. But this reply does not answer the objection for two reasons. First, supervenience is a metaphysically weak relation that can also cover identity theories (Kim, 1998, pp. 10ff.). Second, Chalmers's reply says nothing on the key point of the objection: a criterion of epistemic consistency is not sufficient to state ontological conclusions.

The objection regarding the impossibility of deriving ontological conclusions from premises grounded on the epistemic criterion of internal coherence can also be enforced with a second argument grounded on the following analogy with the history of science.

A century ago a vitalist might have conceived a Chalmers's style argument to support dualistic conclusions on the relations between life and nature. But today we just say that biological and physical properties are different and that the difference relies on the level of description, not on some ontological gap between life and the physical world. Why should the situation be different for consciousness?

Chalmers is aware of the possible parallel between his position and vitalism, but for him, the vitalist argument is invalid because biology logically supervenes on physics, so that this would be an entirely different case with respect to the merely natural supervenience of consciousness on physics. He writes:

I think this reaction misplaces the source of vitalist objections. Vitalism was mostly driven by doubt about whether physical mechanisms could perform all the complex *functions* associated with life . . . But implicit in these very doubts is the conceptual point that when it comes to explaining life, it

is a performance of various functions that needs to be explained. Indeed, it is notable that as physical explanation of the relevant functions gradually appeared, vitalists doubts mostly melted away. With consciousness, by contrast, the problem persists even when the various functions are explained. (1996, p. 109)

But Chalmers's reply appears to miss the point. To understand why, let us see Searle's reply to Chalmers:

In his argument for property dualism, he says, correctly, that you can imagine a world that has the same physical features that our world has—but minus consciousness. Quite so, but in order to imagine such a world, you have to imagine a change in the laws of nature, a change in those laws by which physics and biology cause and realize consciousness. But then . . . if you are allowed to mess around with the laws of nature, you can make the same point about flying pigs. If I am allowed to imagine a change in the laws of nature, then I can imagine the laws of nature changed so pigs can fly. He points out . . . that that would involve a change in the distribution of physical features. . . . But my answer to that . . . is that if consciousness is a physical feature of brains, then the absence of consciousness is *also* a change in the physical features of the world. That is, his argument works to establish property dualism only if it assumes that consciousness is not a physical feature, but that is what the argument was supposed to prove. (1997, pp. 173–174)

Searle's objection is that Chalmers's argument is begging the question. Following this argument, we could also conclude that the concept of logical supervenience and the possibility of the zombie argument are incoherent.

Searle puts forward this further objection and shows first that the argument for dualism is dependent on the descriptions of mental and physical therein used and second, that the argument makes an invalid inference from antireductionism to dualism. The first point states that the difference between logical or natural supervenience is a merely terminological issue:

Whether or not a state of affairs is logically possible depends on how it is described. Is it logically possible that there should be physical particles without any consciousness in the universe? The answer is yes. But is it possible that there should be the trajectories of physical particles as they have in fact occurred together with the laws of nature that, among lots of other things, determine those trajectories to cause and realize consciousness, but minus any consciousness? Then the answer is no. Described in one way, the absence of consciousness is logically possible; described in another way it is not. The picture the dualists have is that the microphysical particles are like tiny grains of sand affected by independent forces, and they can imag-

ine the movement of the sand without any consciousness. But that is a false picture. At the most fundamental level points of mass/energy are constituted by the forces that are described by the laws of nature. From those laws the existence of consciousness follows as a logical consequence, just as does the existence of any other biological phenomena, such as growth, digestion, or reproduction. (Searle, 2004, pp. 129–130)

Kripke's picture described by Chalmers appears to suggest that God creates microparticles first and next regulates the interactions. Under this picture, the argument would follow, but the picture is wrong: the interactions at the microlevel described by natural laws are constitutive of the entities themselves. But then, consciousness is logically implied by the physical in this wider description because a zombie world could not be physically identical to ours. If consciousness is lacking, then the physical interactions determining the existence of consciousness are also lacking.

Therefore, the distinction between logical and natural supervenience is useless because the logical derivability of a state of affairs depends on the description of the state of affairs itself. If we are right on this point, each case of logical supervenience can be described as a case of natural supervenience (and vice versa).

A second thesis supported by Chalmers and criticized by Searle is that antireductionism implies dualism. But antireductionism and dualism are not equivalent:

[T]he dualist supposes that the irreducibility of the consciousness already implies that consciousness is something over and above its neurobiological basis. I deny that implication. . . . The fact that the causal powers of consciousness and the causal powers of its neuronal basis are exactly the same shows that we are not talking about two different things, consciousness and neuronal processes. If two things in the real empirical world have an independent existence, they must have different causal powers. But the causal powers of consciousness are exactly the same as those of the neuronal substrate. This situation is exactly like the causal powers of solid objects and the causal powers of their molecular constituents. We are not talking about two different entities but about the same system at different levels. Consciousness differs from solidity, liquidity, etc. in that its causal reduction does not lead to an ontological reduction. This, as we have seen, is so far an obvious, and indeed trivial, reason. Consciousness has a first-person ontology; neuronal processes have a third-person ontology. For that reason you cannot ontologically reduce the first to the second. Consciousness is thus an aspect of the brain, the aspect that consists of ontologically subjective experiences. But there are not two different metaphysical realms in your skulls, one "physical" and one "mental." Rather, there are just processes going on in your brain and some of them are conscious experiences. (Ibid., pp. 127–128)

In this quotation, we can identify many of the conceptual tools of biological naturalism: the conception of experiences as higher-level brain processes, the distinction between causal and ontological reduction, and the distinction between objective and subjective ontology. With these tools in hand, we can point out the differences between the two approaches (see also Searle, 2002e).

According to Searle, mental states and conscious experiences are higher-level brain processes, multiply realizable (in principle), and ontologically irreducible. For Chalmers, conscious phenomena are ontologically independent processes multiply realized in physical systems according to nomological connections.

The perspectives diverge on this point: according to Chalmers, the irreducibility of consciousness implies ontological independence, while according to Searle, subjective ontology is a part of the biological order of the natural world. One of Searle's arguments for this point maintains the identity of brain and mind causal powers and then argues the causal reducibility of conscious states according to the theses of biological naturalism. Searle concludes that when we talk about consciousness and brain, we are not describing two different things, but just the same system at different levels of its organization.

This argument against dualism is not available for Chalmers because on his account, we could hardly attribute causal powers to consciousness:

Any point of view which takes consciousness seriously will at least have to face up to a limited form of epiphenomenalism. The very fact that experience can be coherently subtracted from any causal account implies that experience is superfluous in the explanation of behavior, whether or not it has some subtle causal relevance. It is possible that it will turn out to be causally irrelevant in a stronger sense. (1996, pp. 158–159)

Perhaps Chalmers does not think of himself as an epiphenomenalist. Yet he thinks that if we accept the merely natural supervenience of consciousness and its non-physical nature, then we should take epiphenomenalism seriously:

On examination, there are not many *arguments* that do serious damage to epiphenomenalism. . . . Arguments aside, some have the *intuition* that epiphenomenalism must be wrong, but the intuition does not suffice to reject the position in the face of strong arguments in its favor. (Ibid., p. 160)

The problem is not so much that epiphenomenalism sounds “repugnant,” as he also writes in another point of the book, but that his arguments are far from compelling. In Chalmers's perspective, we can say that the problem of mental causation is more the product of question-begging arguments (naturalistic property dualism and zombie argument) instead of the product of an independently valid reasoning.

5. Desubstantialization of Mind, Functionalism, and Conceptual Dualism

We examined the main functionalistic theories of consciousness (Dennett's eliminativism and Chalmers's neodualism) to argue two of the main theses of the present work. First, the desubstantialization of mind is insufficient to define a coherent theory of the ontology of consciousness. Second, we have to integrate the desubstantialization with a new naturalistic framework to overcome the traditional monism-dualism dilemma.

Here we have to focus on two theoretical presuppositions (the desubstantialization of mind in its functionalistic interpretation and conceptual dualism) shared by these authors to see that Dennett and Chalmers are part of a tradition that, although maybe still dominant, does not give good answers to the problem of consciousness.

Dennett and Chalmers develop their analyses from the premise that "mind" does not name a thing, but a dynamic process multiply realizable and causally responsible for the management of organism-environment relations.

Their analyses diverge when they try to satisfy Chalmers's first adequacy condition: taking consciousness seriously.

Dennett thinks that under some empirical (the parallel and "competitive" brain architecture), philosophical (the concept of quale is contradictory), and epistemic (objective and reductionist science vs. subjective consciousness) presuppositions, the elimination of the qualitative and subjective dimension of mind follows.

The problem with this position is not so much that it is absurd: the history of philosophical and scientific thought already accustomed us to accept initially absurd and counterintuitive positions. The problem is that Dennett does not state sufficient reasons to justify this conclusion. His eliminativism works only once we interpret phenomenal consciousness according to the discredited Myth of the Cartesian Theater. But the argument is at best tendentious because it works only if we confuse the (ontological) subjectivity of mind with (epistemic) certainty. The refutation of the Myth is not sufficient to deny the existence of consciousness as a phenomenon for which a scientific or philosophical theory has to account. But we have to remember that "to account for" is not the same as "reducing" or "eliminating" a phenomenon, as Searle's analyses of reduction and bottom-up no time gap explanation showed.

Chalmers, instead, thinks that taking consciousness seriously implies not only the failure of materialism but also the non-physical nature of consciousness and a dualistic ontology. We could state the bridge laws governing psychophysical relation, but we will never know why our world, which includes conscious experience, evolved in a totally different way with respect to the hypothetical zombie world.

Chalmers's property dualism does not follow for reasons analogous to those adduced against Descartes' dualistic argument: He makes an invalid inference from epistemic premises to ontological conclusions and does not take into account the historical and cultural constraints acting on logical conceivability. If

this objection is correct, then we can prove dualism only with an argument that begs the question.

Dennett and Chalmers explicitly agree on another point: their acceptance of conceptual dualism. Dennett (1991, p. 37) declares that materialism is the contemporary orthodoxy because accepting dualism is like giving up on the task of understanding mind. The concept of a mental substance or property ontologically independent of the physical gives rise to unnecessary problems and it is intrinsically self-contradictory, but since mental substances do not exist, we are simply forced to accept some kind of materialism: no *res cogitans*, no consciousness (Dennett, 1997).

Chalmers is even more explicit on this issue:

The failure of materialism leads to a kind of dualism: there are both physical and non-physical features of the world. The falsity of logical supervenience implies that experience is fundamentally different in kind from any physical feature. (1996, p. 124)

Once Dennett argues that dualism fails, he does not see any options other than eliminativism, the counterintuitive results of which are not well supported by his arguments. On the other hand, even if we take for granted the validity of Chalmers's arguments, we are left with a mystery: how is it possible that conscious beings are part of the natural world? Since consciousness is non-physical, how could it make a difference in a causally closed physical world? Chalmers does not go beyond an odd "neutrality" on the issue of mental causation (1997; Chalmers and Searle, 1997).

But the inference from the failure of materialism to dualism is correct, from Chalmers's point of view, because the argument implicitly assumes that consciousness cannot, as subjective, qualitative, unified, and irreducible, be part of the natural world.

The argument of this chapter has produced the following conclusion. The main functionalistic theories of consciousness share, beyond desubstantialization, the same implicit premise of Kim's causal exclusion dilemma: mind per se is ontologically incompatible with the physical per se. Do the troubles with these theses derive from conceptual dualism?

This is Searle's diagnosis of the problem and he reconstructs a set of principles deriving from conceptual dualism that makes reductionism and eliminativism appear to be the only viable solutions to the mind-body problem. Materialism appears to be a solution in light of the contradictory nature of dualism but, at the same time, the principles make it impossible to conceive of the conscious mind as part of the natural world. These principles are:

- (1) Where the scientific study of the mind is concerned, consciousness and its special features are of rather minor importance. It is quite possible, indeed desirable, to give an account of language, cognition, and mental

states in general without taking into account consciousness and subjectivity.

(2) Science is objective. It is objective not only in the sense that it strives to reach conclusions that are independent of personal biases and points of view, but more importantly, it concerns a reality that is objective. Science is objective because reality itself is objective.

(3) Because reality is objective, the method in the study of the mind is to adopt the objective or third-person point of view. . . .

(4) From the third-person, objective point of view, the only answer to the epistemological question “How would we know about the mental phenomena of another system?” is: We know by observing its *behavior*. . . .

(5) Intelligent behavior and causal relations to intelligent behavior are in some way the essence of the mental.

(6) Every fact in the universe is in principle knowable by human investigators. Because reality is physical, and because science concerns the investigation of physical reality, and because there are no limits on what we can know of physical reality, it follows that all of the facts in the universe are knowable and understandable by us.

(7) The only things that exist are ultimately physical, as the physical is traditionally conceived—as opposed to the mental (Searle, 1992, pp. 10–11).

If these analyses are correct, then we have a way to make the essential features of consciousness compatible with a naturalistic framework and, within this framework, with the desubstantialization of mind.

Biological naturalism gives us a way to get rid of conceptual dualism and to refute the principles underlying a debate that Searle compares with the behavior of a neurotic person, whose neurosis “takes the form of repeating the same pattern of behavior over and over” (*ibid.*, p. 31).

Once this is done, we have to understand mental processes (as mental *and* as processes) as a part of animal biology.

The subjectivity of consciousness is not an obstacle to a scientific understanding of mind. Conversely, the possibility of a scientific explanation of mind will not imply reductive procedures that, leaving out its essential features, would lead to its elimination.

The critique of conceptual dualism and the generality claim justifying biological naturalism are the reasons that allow Searle’s theory to be immune from Dennett and Chalmers’s problems. Here we can find a model for a coherent conjugation of naturalism and antireductionism. But how can we argue for this interpretation in a constructive, positive way?

I outlined the problems of reductionism and eliminativism, but these philosophies have had new life during the last decade, mainly because of the difficulty encountered by antireductionism in dealing with mental causation, which many philosophers, such as Dennett and Kim, see as internal to all non-reductive theories in general and to Searle’s view especially.

Dennett (1993; 2005) argues that while his eliminativism can deal with mental causation, the antireductionism of biological naturalism forces Searle to embrace epiphenomenalism.

But Dennett's interpretation not only does not take into account the debate between Searle and Kim, but also some core concepts and theses of Searle's theory that I will deepen in the present work.

- (1) Action and perception are biologically and logically primitive mental states in virtue of their power to causally and intentionally relate organism and environment.
- (2) Action and perception are interdependent.
- (3) Memory, perception, and intentional action are causally self-referential: being in causal relations is not an external event with respect to the contents of those states; instead it is an internal feature of their logical structure.

Di Lorenzo Ajello noticed that action is the paradigmatic model of intentional states in Searle's theory and that the concept of causal self-referentiality grounds this thesis:

In Searle's theory returns an image of mind, and therefore of human beings' acting and speaking, focused on a concept of Intentionality intrinsically connected with causality, in which it assumes a paradigmatic value, the idea of intentional action as action that is such because there is nothing accidental in the realization of the plan of mind underlying it. . . . But it is through the theory of causal self-referentiality of intentions . . . perceptions, and memories, that such a paradigmaticity reveals its crucial role: in this theory there is the peculiarity of intentional action as such (my translation). (2001, pp. 86–87)

Also, Dennett's interpretation does not take into account that in Searle's opinion, dualism cannot give an intelligible account of mental causation and that this inability would make dualism as a fundamentally incoherent position:

There really is a distinction between those irreducible features of the world that have a first-person ontology and those that do not. But it is a deep mistake to suppose that that real distinction is the same as the old-time distinction between the mental and the physical, between *res cogitans* and *res extensa*, or that the subjective phenomena are something over and above the systems in which they are realized.

The dualist thinks "irreducibility" already implies that the irreducible phenomenon is something over and above its physical basis. This poses an impossible problem for the property dualist: either consciousness functions causally or it does not. If it does, then we appear to have causal overdeter-

mination: if I intentionally raise my arm it appears that my arm going up has two causes, one physical, one mental. But if consciousness does not function causally then we have epiphenomenalism. No such problem arises for biological naturalism, because the causal functioning of consciousness is just a form of brain functioning described at a level higher than neurons and synapses. Think of it this way: roughly speaking, consciousness is to neurons as the solidity of piston is to the metal molecules. Both consciousness and solidity function causally. But neither is over and above the system of which they are a part. (Searle, 2004, pp. 130–131)

Searle claims that his approach can efficaciously solve the problem and grounds his claim on an analogy between consciousness and other emergent properties. But we can also say that this solution is efficacious because we can generalize it through the dissolution of conceptual dualism and the development of a detailed model of the ontology of mind grounded on a non-reductive view of reality and scientific work.

I began to analyze this issue to show that biological naturalism can give good solutions to the problems of mental causation and consciousness, not only by virtue of the critique of conceptual dualism, but also through a general explanatory model of psychophysical relation. In the next chapters, I will argue that Searle's theory of intentionality and the thesis of the holistic structure of mind, worked out within the framework of biological naturalism, can face other problems, such as the problems of the intersubjective structure of consciousness and the issue of the emergence of a self-conscious and rational self.

I will compare Searle's holistic model with some of the most recent achievements in neurosciences, especially with Edelman's and Tononi's theory of consciousness and with Antonio Damasio's theory of the self. I will examine whether biological naturalism can provide a useful conceptual framework to interpret and understand current scientific research and, conversely, whether modern research in the neurosciences can offer some confirmation of Searle's theory of biological naturalism.

Four

HOLISM AND MENTAL CAUSATION IN THE THEORY OF INTENTIONALITY

The historical development of philosophy of mind went, in modern philosophy, from Cartesian substantialism to the desubstantialization of mind. Philosophers and scientists began to see mind as a set of irreducible processes driving the interactions with the environment. They tried to overcome the traditional monism-dualism dilemma, and conceived mind as part of the biological make-up of living beings.

But the traditional monism-dualism dilemma returned, even within the desubstantialization of mind in its contemporary version (functionalism), giving rise to the return of the traditional problems of consciousness and mental causation.

Daniel C. Dennett and David J. Chalmers's theories of consciousness, and Jaegwon Kim's causal exclusion dilemma are examples of the traditional conceptual pattern with its related problems. On the materialist side, philosophers offer models the conclusions of which eliminate the subjective dimension of living creatures like us and other non-human animals from the scope of the natural world. On the dualistic side, philosophers transform the existence in nature of this subjective aspect into a mystery.

John R. Searle construes conceptual dualism as the premise underlying these problems. The implicit acceptance of a mutual ontological exclusion between the mental and the physical shapes our discussions of the mind-body problem determining its dilemmatic structure.

On the contrary, Searle's biological naturalism starts with the critique of conceptual dualism, justifies its non-reductive embodiment of mind with an analysis of the explanatory structure of natural sciences, and gives a version of the desubstantialization of mind that is immune from the problems of positions like those worked out by Dennett, Chalmers, and Kim.

My analysis of biological naturalism outlines some core theses of Searle's approach. First, we have the logical interdependence between subjectivity, qualitateness, and holistic structure of the unified field of consciousness: We cannot have qualitative states if these states are not subjective (they exist only because the subject having experiences exists) and we cannot have this qualitative subjectivity if it does not structure its states in a unified field. Every state has its identity only within a unified field of other mental states.

Second, the non-reductive embodiment of mind involves the identification of a causal-evolutionary role for mind and consciousness, which Searle sees as systems devoted to the creative structure of the relationships between organism and environment (Searle, 1992, pp. 107–109).

In sum Searle's theory of intentionality provides a ground for these theses through the concepts of Network, Background, and causal Self-referentiality. Next, I will discuss the objection that Searle's theory of intentionality would imply solipsism, which is the position that the structure and existence of mental states is independent of the embodiment of the agent in a body and in a physical and (for human beings) cultural environment. To answer this critique, I will analyze the extension of Searle's theory to collective actions and I will compare the results of the analysis with the discovery of mirror neurons.

The analysis of these concepts will prepare the ground for the critical comparison between Searle's philosophy of mind and contemporary research in neurosciences. I will just sketch these ideas here and develop them in more detail in the next chapter.

1. The Logical Structure of Intentional Representations: Psychological Mode, Representational Content, Direction of Fit, and Conditions of Satisfaction

The analysis of the logical structure underlying the functioning of intentional representations is a crucial step in Searle's philosophy of mind for which we can offer two main reasons. The analysis grounds Searle's speech acts theory, worked out in *Speech Acts* (1969) and *Expression and Meaning* (1979a). In these essays, Searle states two theses concerning the functioning of language, which commit the philosopher to give an account of the intentionality of mind underlying the performance of speech acts.

First, every performance of a speech act commits the speaker to a sincerity condition with respect to the mental state expressed by the speech act (Searle, 1969, pp. 60, 65).

Second, we can give a complete analysis of the ways we can do things with words through a taxonomy of five classes of illocutionary acts: assertives (we commit ourselves to tell how the world is), directives (we tell others how to change the world), commissives (we commit ourselves how to change the world), expressives (we express our psychological state), and declarations (we create a new reality saying that it exists). The reason is that the classes express the more fundamental ways through which mind can represent the world (Searle, 1979b).

In addition, according to Searle, the structure of intentionality is the structure of all conscious life (cf. Searle, 2004, p. 174).

To these two main reasons, we can add a third: Conscious mental representations have a causal role (the management of the organism-environment relations) and the analysis of the formal structure of intentionality reconstructs the mechanisms allowing these representations to exercise their role.

The development of a theory requires a conceptual apparatus, which in the case of intentionality, is borrowed from the theory of speech acts. We are talking about the concepts of direction of fit, conditions of satisfaction, and the notions of mode and content of a mental state. Taken together, these concepts define the

notion of representation and describe the concept of intentional state as the representation of its conditions of satisfaction:

To say that a belief is a representation is simply to say that it has a propositional content and a psychological mode, that its propositional content determines a set of conditions of satisfaction under certain aspects, that the psychological mode determines a direction of fit of its propositional content, in a way that all of these notions . . . are explained by the theory of speech acts . . . the formal relations between these various notions can be stated as follows: every Intentional state consists of an *Intentional content* in a *psychological mode*. Where that content is a whole proposition and where there is a direction of fit, the Intentional content determines the *conditions of satisfaction*. Conditions of satisfaction are those conditions, as determined by the Intentional content, that must obtain if the state is to be satisfied. For this reason the *specification* of the content is already a *specification* of the conditions of satisfaction. . . . Since all representations . . . are *always* under certain aspects and not others, the conditions of satisfaction are represented under certain aspects. (Searle, 1983, pp. 12–13)

We can see first the distinction between mode and content. So for example, Searle writes, I can ask that you leave the room, I can order that you leave the room, I can predict that you will leave the room, and so on. When I speak, I can use the same content with many illocutionary forces deriving from the five fundamental classes and the same holds for intentionality. I can believe, hope, desire, or fear that you leave the room.

So speech acts and intentional states have the same structure that we can formalize, for intentional states, as $S(p)$, where S symbolizes the type of state, or psychological mode, and p stands for the representational content, that we can express in the propositional form or as having just an object as its content (we can believe that snow is white or we can love Sally).

Psychological mode (like illocutionary force) determines the direction of fit. Some intentional states, like beliefs, are satisfied when their contents fit an independently existing world, while others, like desires, are satisfied when the world changes so that it fits what is expressed in the contents.

So Searle writes:

Intuitively we might say the idea of direction of fit is that of responsibility for fitting. If the statement is false, it is the fault of the statement (word-to-world direction of fit). If the promise is broken, it is the fault of the promiser (world-to-word direction of fit). . . . Now something very much like these distinctions carries over to Intentional states. If my beliefs turn out to be wrong, it is beliefs and not the world which is at fault, as is shown by the fact that I can correct the situation simply by changing my beliefs. It is the responsibility of the belief, so to speak, to match the world, and where the

match fails I repair the situation by changing the belief. But if I fail to carry out my intentions or if my desires are unfulfilled I cannot in that way correct the situation by simply changing the intention or desire. In these cases it is, so to speak, the fault of the world if it fails to match the intention or desire, and I cannot fix things up by saying it was a mistaken intention or desire in the way I fix things up by saying it was a mistaken belief. (Ibid., pp. 7–8)

Searle traces the genesis of the concept back to John L. Austin and Elizabeth Anscombe, although this notion and the associated terminology also refer to an evolutionary theoretical context (Di Lorenzo Ajello, 2001, pp. 76–79).

As Searle writes elsewhere (1998, p. 99), we can compare intentionality with the action of throwing an arrow toward a target. In both cases, success depends on satisfying some conditions. Intentional states determine the target through the content and with different directions of fit as determined by psychological modes, so that success for each state is to be assessed in relation to the whole structure of the state (mode, content, and direction of fit).

Like speech acts, intentional states determine the conditions under which they succeed or fail: a belief is satisfied (true) if and only if the world is as represented in the content, while a desire is satisfied if and only if the world changes so that it fits the way it is represented in the content.

Since the conditions of satisfaction specified by the content are determined as a function of mode, content, and direction of fit, the analysis of conditions of satisfaction of each intentional state is the key point to understand the underlying logical structure.

Searle also says that every intentional state always represents its conditions of satisfaction under some aspects and not others:

My conscious experiences, unlike the objects of the experiences, are always perspectival. They are always from a point of view. But the objects themselves have no point of view. Perspective and point of view are most obvious for vision, but of course they are features of our other sensory experiences as well. . . .

Noticing the perspectival character of conscious experience is a good way to remind us that *all intentionality is aspectual*. Seeing an object from a point of view, for example, is seeing it under certain aspects and not others. In this sense, all seeing is “seeing as.” And what goes for seeing goes for all forms of intentionality, conscious and unconscious. All representations represent their objects, or their conditions of satisfaction, under aspects. Every intentional state has what I call *an aspectual shape*. (Searle, 1992, p. 131)

Conscious experiences and intentional states have an aspectual shape or perspectival element and this holds not only for perceptual experiences, but also for the complex cognitive and volitional states like beliefs and desires. We can

believe that Clark Kent is a reporter but that Superman is not, even though Kent and Superman are the same person; we can analogously desire to drink a glass of water but not a glass of H₂O.

Gottlob Frege (1980) analyzed this point only in relation to language, without considering mind and mental states. He distinguished between the *Sinn* (Sense) expressed by a name or sentence and the *Bedeutung* (reference), where the first is the aspect under which we are talking about something and the second is the thing to which our talking is directed. Searle's distinction between the aspectual shape of mental states and the object or state of affairs to which mind is directed generalizes Frege's analysis and has its same conceptual point: the mind-world relationship has an aspectual structure. We never access the world *per se*, independently of our cognitive structures. Instead, whenever we talk or think about something we represent it under aspects. These are real features of the reference to which we are directed, but the way we determine these aspects depends on our cognitive structures (Di Lorenzo Ajello, 2001, pp. 80–81). But Searle goes far beyond Frege's philosophy of language, since the aspects (senses) under which the reference falls, are determined, in Searle's theory of intentionality, holistically, pragmatically, and causally. The next sections will introduce the concepts which allow us to analyze content determination this way: the holistic Network of intentional states, the pragmatic Background of preintentional skills, and the causal Self-referentiality of perception and action.

2. The Causal Self-Referentiality of Perception and Action

One of the main features of Searle's theory of intentionality is that the results of his analyses lead to the thesis that perception and action, instead of beliefs and desires, must be considered as paradigmatic intentional states (Lepore and Van Gulick, 1991b, p. xii; Di Lorenzo Ajello, 2001, pp. 79–87, 90).

At first glance the formal features of contents of visual experiences and intentions are fully analogous to beliefs and desires. A perceptual experience has a propositional content (even though I see a yellow station wagon, what I am seeing is *that* the yellow station wagon is in front of me) and mind-to-world direction of fit (the visual experience is responsible for fitting an independently existing reality). An intention, like a desire, has a propositional content and world-to-mind direction of fit. But the intentional contents of perception and action structure the internal logical relations with their respective conditions of satisfaction in a more complex way. As Searle notices (1983, p. 81), we have a clue of this more detailed internal structure: intentions are the only intentional states whose conditions of satisfaction have a name, "action." Nothing similar happens with beliefs and desires.

In perception, the intentional content, when satisfied, does not merely *represent* a state of affairs as its condition of satisfaction. Instead, perception *presents* it directly, although always under determined aspects. The verb "seeing," unlike "believing," is a success-verb: I see something only if what I am seeing

exists and it is causing my visual experience (*ibid.*, pp. 41n, 49). If this is not the case, then maybe I am hallucinating. But, Searle thinks, I am *seeing* nothing. To say that even in cases of hallucination we see something (the experience itself) is typical of representational theories of perception that Searle (2004, pp. 260–261) attributes to the philosophical tradition ranging from René Descartes to Immanuel Kant. This conception would depend on the categorical mistake of attributing to visual experiences features belonging to their conditions of satisfaction. So if I say that my experience of the yellow station wagon is itself yellow, my sentence is as meaningless as to say that it has a power of 1,300 cc. (Searle, 1983, pp. 43–45). Searle also showed, through a transcendental argument, that normal understanding of a wide class of speech acts made in a public language crucially depends on speakers' ability to perceptually access the real world (2004, pp. 273–277). "Believing," on the contrary, is not a success-verb, since we can believe that Santa Claus exists even though he does not, in reality, exist.

That "seeing" is a success-verb suggests that this difference has to be reflected in the contents of the mentioned mental states. The difference emerges when we consider that while for a belief, the *way* we get satisfaction and mind-to-world fitness is irrelevant, the same does not hold for perception. So if I believe, based on a mistaken argument, that the Pythagorean theorem is valid, we could say that my belief is not well justified, but not that my belief is not true. In addition, we could never say that I do not really have that belief. But if, while I am looking at my coffee, the resulting visual experience is not caused by the coffee itself, but, say, by its smell, then I am not having a visual experience of the coffee at all: what I am having is just an hallucination.

This asymmetry in the structures of the contents of belief and perceptual experience allows Searle to introduce the concept of causality as a logical feature of the perceptual content. He expresses this point saying that the content of perception is causally self-referential:

The content of the visual experience is . . . self-referential . . . what the Intentional content requires is not simply that there be a state of affairs in the world, but rather that the state of affairs in the world must cause the very visual experience which is the embodiment or realization of the Intentional content. . . . The Intentional content of the visual experience therefore has to be made explicit in the following form:

I have a visual experience (that there is a yellow station wagon there and that there is a yellow station wagon there is causing this visual experience). (1983, p. 48)

Therefore, perception is a causal and intentional relationship between mind and world. Searle integrates this feature within the more general theory of intentional states as follows:

On this account perception is an Intentional and causal transaction between mind and world. The direction of fit is mind-to-world, the direction of causation is world-to-mind; and they are not independent, for fit is achieved only if the fit is caused by the other term of the relation of fitting, namely the state of affairs perceived. (Ibid., p. 49)

The internal relationship among intentionality, causality, and conditions of satisfaction is perhaps even more evident in the case of intentions. Like perception, but unlike desire, intentions have a causally self-referential element as a constitutive part of their contents.

Donald Davidson, Roderick M. Chisholm, and Jonathan Bennett worked out some arguments against the idea that the relation between intention and action could be causal. Suppose that I have the intention to kill my uncle by means of hitting him with my car. The intention will not be satisfied if I hit him because the homicide intention produces in me such nervousness that before I intentionally strike him, I cause a car accident in the course of which my uncle happens to die (ibid., pp. 81–83, 107–111). So we still have to qualify the causal relation between intention and action if it has to hold.

Unlike in the case of desires, to satisfy an intention, it is not sufficient that the conditions of satisfaction just happen. It would still be insufficient if the causal relation between intention and action were external to the content of the intention itself, like in the case of my nervousness causing the unintentional car accident. An intention has to be carried out, where “to be carried out” means that the intention is satisfied if and only if the conditions of satisfaction occur as effect of the intention itself and this relation is represented in the content. Like perception, the content of intention has a causally self-referential element:

Both prior intentions and intentions in action are causally self-referential in the same sense that perceptual experiences and memories are causally self-referential. That is, like perceptual experiences and memories their conditions of satisfaction require that the Intentional states themselves stand in certain causal relations to the rest of their conditions of satisfaction. . . . Suppose I intend to raise my arm. . . . The content of my intention can’t be that my arm goes up, for my arm can go up without me raising my arm. Nor can it be simply that my intention causes my arm to go up, for . . . a prior intention can cause a state of affairs represented by the intention without that state of affairs being the action that would satisfy the intention. . . . The Intentional content of my intention must be at least (that I perform the action of raising my arm by way of carrying out *this intention*).

But what is meant by “carrying out” in this formulation? At least this much: If I am carrying out that intention then the intention must play a causal role in the action, and the argument for this is simply that if we break the causal

connection between intention and action we no longer have a case of carrying out that intention. (*Ibid.*, pp. 85–86)

Causal self-referentiality justifies Searle's thesis that intentions and visual experiences are irreducible to sets of beliefs and desires and that they are, instead, primitive mental states:

[I] believe that the power and scope of an approach to Intentionality in terms of conditions of satisfaction will become more apparent as we turn . . . to what I take to be the biologically primary forms of Intentionality, perception and action. Their Intentional contents differ from beliefs and desires in a crucial respect: they have Intentional causation in their conditions of satisfaction. . . . Beliefs and desires are not the primary forms, rather they are etiolated forms of more primordial experiences in perceiving and doing. Intention, for example, is not a fancy form of desire; it would be more accurate to think of desire as a faded form of intention; intention with the Intentional causation bleached out. (*Ibid.*, p. 36)

In Searle's theory, actions and perceptions assume a more crucial role than is usually allowed in the abstract theoretical models of Artificial Intelligence because they are characterized as paradigmatic and primitive intentional states in the logical and biological sense, which give rise to a fruitful interaction with linguistic abilities and other higher-level mental processes. These interactions structure more complex forms of mental processes and abilities, until they produce intentional states that would be impossible without those higher capacities:

The picture I have is this: a human child begins with prelinguistic forms of Intentionality. By a kind of bootstrapping effect the child acquires primitive linguistic expressions of that Intentionality. But a little bit of language goes a long way; and the child develops a richer Intentionality which it could not have developed without linguistic forms. This richer Intentionality enables a further richer linguistic development, which in turn enables richer Intentionality. All the way up to the developed adult, there is a complex series of developmental and logical interactions between Intentionality and language. Most forms of adult Intentionality are essentially linguistic. But the whole edifice rests on biologically primitive forms of prelinguistic Intentionality. (Searle, 1991a, p. 94)

"Bootstrapping effect" has here a technical meaning, as used in neuroscientific literature. Gerald M. Edelman (1989), for example, puts forward the hypothesis that the complex and rich syntax of natural languages emerges only after a "semantic bootstrapping," the process of self-elevation through which the brain autonomously acquires complex semantic abilities without environmental "instructions" or internal homunculi. This self-elevation process happens at the phy-

logenic level in the course of evolution and at the ontogenetic level in interactions with the environment.

Perception in adult human beings is a paradigmatic example of this fruitful interaction between primitive and higher-level cognitive processes. But Searle's theory of causal self-referentiality apparently implies a causal theory of perception of a behaviorist kind: the environmental stimulus causes the visual experience (Freeman, Skarda, 1991, p. 125). But we know from the Gestalt Theory, from Jean Piaget's studies in genetic epistemology (1972; 1975), and from the studies in perceptual categorization (Edelman, 1989; 1992), that the stimulus will be never be sufficient *per se* to cause perception. We always "pre-understand" a stimulus in a Network of action schemes and perceptual categories, we always see something "as" something.

Some words in natural languages are ambiguous because they have multiple meanings and numerous connotations. Pictures like Ludwig Wittgenstein's duck-rabbit illusion and other cases of ambiguous perception show that perceptual intentional contents have, in some cases, the same sort of ambiguity of meaning. But the phenomenon is quite general, so that a great part of our experiences would not be possible without the existence of a set of mental abilities among which linguistic skills play a crucial role. Exceptions to this point do exist: we *see* the moon or the sun as "bigger" when it changes its position in the sky although we believe that the physical dimensions of these objects remain the same over time regardless of relative position. We also have the phenomenon of Ernst Mach's bands: we see a uniformly grey band as "lighter" when placed near another, darker band, than when we view it alone (Crane, 2003, pp. 149–150).

But exceptions of this kind are borderline cases that do not obscure the interesting phenomenon of the embedding of mental states in a complex set of other mental phenomena and abilities.

To understand in detail how Searle's theory can account for these highly complex and articulated structures I will expand upon the concepts of Network and Background.

3. Network and Background: The Holistic and Pragmatic Structure of Mind

The analysis so far does not take into account the complications in the determination of content introduced by what we could call its context-dependence (Fotion, 2000, pp. 117–127). By "context," I mean the embedding of every single state in a representational Network made of other intentional states, and as the operation of the Network itself against a wider system of nonintentional abilities and capacities composed of "deep" biologically grounded skills and social, cultural, or "local" practices.

Here is the formulation of this thesis:

- (1) Intentional states do not function autonomously. They do not determine conditions of satisfaction in isolation.
- (2) Each intentional state requires a Network of other intentional states. Conditions of satisfaction are determined only relative to the Network.
- (3) Even the Network is not enough. The Network only functions relative to a set of Background capacities.
- (4) These capacities are not and cannot be treated as more intentional states or as a part of the content of any particular intentional state.
- (5) The same intentional content can determine different conditions of satisfaction (such as truth conditions) relative to different Backgrounds, and relative to some Backgrounds it determines none at all. (Searle, 1992, p. 177)

As Searle notes, the context-dependence of linguistic and mental representations is, by itself, a part of contemporary philosophical orthodoxy. From Wilfrid Sellars to Davidson and Robert B. Brandom, philosophers think that the representational Network of intentional states is articulated by inferential constraints. This means, among other things, that rationality would be constitutive of mental phenomena so that the difference between non-human and human mental phenomena (such as rational thought) would not be a matter of degrees. The holistic structure of mind is governed by the validity of inferential constraints and, ultimately, by validity standards realized in social practices to which only a language shared in a community can give access.

From these premises, Davidson (1984b) argues his famous but disputed thesis that nonlinguistic animals lack mental phenomena, while John McDowell (1994), along the same lines, says that nonlinguistic animals lack a defined subjectivity because they do not have a conceptually shaped orientation toward the world.

Even Searle argues that the constraints of rationality are not external with respect to mental phenomena. Instead, these constraints would be internally construed in the logical structure of states such as beliefs and desires:

Being subject to rational criteria of assessment is internal to and constitutive of intentional phenomena, in a way that winning and losing are constitutive of football games. You don't first have beliefs, hopes, desires, and intentions, and then external to them introduce rational forms of assessment; rather to have the beliefs, etc, is already to have phenomena that are subject to these norms. (2001, p. 109)

Notwithstanding these considerations, the thesis that inferential constraints exhaust the structure of the Network is, according to Searle, mistaken for two reasons. First, this conception of the holistic Network would imply, in Davidson's interpretation of it, dependence of mind on language that would not allow

us to attribute minds to prelinguistic animals and children. At the same time, Searle aims at explaining language on the basis of a biologically primitive and prelinguistic intentionality (1983, pp. 5–13; 2006). In addition, this language-dependence of mental states appears to be independently refuted by experiments on heavily impaired subjects with brain damages in language areas but with all the rest of their mental abilities untouched (Pinker, 1994; Damasio, 1999, chap. 7).

But how do we reconcile Searle's critique of the orthodox position with the intrinsically normative nature of intentional states?

Searle recently clarified the issue analyzing the relation between a "primitive" normativity that articulates the structure of intentional phenomena in prelinguistic animals, and the ability of human beings to publicly undertake commitments in the performance of speech acts:

[I]f you think about matters from the point of view of sweaty biological beasts like ourselves, normativity is pretty much anywhere. The world indeed consists of facts that are largely independent of us, but once you start representing those facts, with either direction of fit, you already have norms, and those norms are binding on the agent. All intentionality has a normative structure. If an animal has a belief, the belief is subject to the norms of truth, rationality, and consistency. If an animal has intentions, those intentions can succeed or fail. If an animal has perceptions, those perceptions either succeed or fail in giving it accurate information about the world. . . . From the point of view of the animal, there is no escape from normativity. The bare representation of an *is* gives the animal an *ought*. What is special about human animals is not normativity, but rather the human ability to create, through the use of language, a *public* set of commitments. (2001, p. 183)

The difference between prelinguistic and linguistic phenomena, therefore, is not a difference in kind, but in degree. In Searle's opinion, we can say that mental phenomena have a prelinguistic normativity that becomes public in typically human rational phenomena when normativity is also enforced by the complex syntax of natural languages (Searle, 2006).

Searle also states a second reason for the incompleteness of the inferentialist analysis: the Network determines conditions of satisfaction only working against a Background of nonrepresentational mental capacities that are neither part of the intentional content nor can be treated as intentional states. The "working against" metaphor does not mean that the Background is external with respect to the operations of the Network:

The Background is not on the *periphery* of Intentionality but *permeates* the entire Network of Intentional states; since without the Background the states could not function, they could not determine conditions of satisfac-

tion. Without the Background there could be no perception, action or memory (Searle, 1983, p. 151)

Here is the hypothesis of the Background sketched in some more detail:

The Background is a set of nonrepresentational mental capacities that enable all representing to take place. Intentional states only have the conditions of satisfaction that they do, and they only are the states that they are, against a Background of abilities that are not themselves Intentional states. In order that I can now have the Intentional states that I do I must have certain kinds of know-how: I must know how to do things, but the kinds of “know-how” in question are not, in these cases, forms of “knowing that.” (Ibid., p. 143)

Searle worked out the concept of Background as a thesis concerning the interpretation and understanding of literal meaning (1979c; 1980) and in *Intentionality*, he extended its scope to the more fundamental forms of prelinguistic intentionality.

This notion turns over the traditional rationalist thesis of the primacy of “knowing that” on “knowing how” (Sellars, 1963). It also criticizes the cognitivist thesis that higher-level cognitive capacities and intelligent behavior depend on gradually “internalized” sets of explicit rules. Searle thinks of his position as controversial, presumably in relation to this classical computational position. Dennett, for example, proposed Ray Jackendoff’s model (articulating computations at the higher-level in lower-level sub-computations) as a clearer alternative with respect to Searle’s thesis, but Dennett appears to have in mind the possibility of escaping the difficulties of classical computational models with his homuncular functionalism (1969; cf. Armstrong, 1991; Searle, 1991c).

To understand why we should accept the thesis of the Background instead of the computational model, consider the following argument:

Suppose that the computational model is right and mind is a system of computations applied to data structures. Suppose also that we program a machine to play chess. For every possible move, the machine will compute many alternatives, each one in turn producing an enormous number of possible consequent moves generating the so-called problem of combinatorial explosion: every step produces a growing number of computations.

We can solve the problem either by “brute force” (constructing a machine that can process millions of computations per second) or “heuristics” (introducing explicit rules allowing the machine to discard less probable game strategies). But possibilities and interpretive problems are virtually endless for the following reasons:

First, many empirical data confirm that our brains are much too slow to make the millions of computations per second required by the brute force strategies: no human being has the computational power of Deep Blue, the famous

computer that won against Garry Kasparov (world chess champion) during 1996 and 1997 (Dennett, 1987, pp. 323–337; Llinàs, 2001).

Second, an indefinitely wide list of algorithmic subprocedures (heuristics) produces a regress of rules to interpret rules.

To illustrate the problem (also known as the “frame problem”) in a more vivid way, consider the following example (Crane 2003, pp. 118–119): Suppose a bus driver is not allowed to change the route of his vehicle unless his company explicitly authorizes him to do so via radio. Suppose also that while the radio is out of order, a passenger suffers a cardiac infarct. The computational model predicts that the driver will not change the route without authorization, even though this kind of behavior does not exemplify a model of intelligence. So we introduce some heuristics such as, “In case of a health emergency, change the route and go to the hospital. Do not wait for authorization.”

But we have further problems of interpretation. We still have to define what counts as a “health emergency.” Is a cold a health emergency? Which hospital should the bus driver go to? To the nearest hospital? Even in case of a terrorist attack? Even in case of some kind of natural catastrophe? We would still have to define what counts as a natural catastrophe and terrorist attack.

Searle thinks that these problems do not arise if we say that we have an implicit know-how, a set of Background preintentional capacities enabling intentional states to work. On this approach:

The rules do not become “wired in” as unconscious Intentional contents, but the repeated experiences create physical capacities, presumably realized as neural pathways, that make the rules simply irrelevant. (1983, p. 150)

He then shows how we can explain the data concerning the sophisticated abilities and competencies of, for example, the expert skier, as compared with the less sophisticated beginner:

with a more economical explanatory apparatus if we do not have to suppose that each physical skill is underlain by a large number of unconscious mental representations, but rather that repeated practice and training in a variety of situations eventually makes the causal functioning of representation unnecessary in the exercise of the skill. The advanced skier doesn’t follow the rules better, rather he skis in a different sort of way altogether. (Ibid., p. 151)

So we could have an alternative explanation of the more flexible, creative fitness of the expert. The explanation would not be that we have a set of unconscious set of explicit rules and heuristics with an infinite regress, but just:

[T]he body takes over, and the skier’s Intentionality is concentrated on winning the race. This is not to deny that there are forms of Intentionality

involved in the exercise of skills, nor is it to deny that some of this Intentionality is unconscious. (Ibid.)

In the end, the pragmatic and holistic nature of intentionality and consciousness takes into account the failure of behaviorism as the correct approach to understand human behavior and mental states.

Searle asserts that the answer is based on his at once holistic, pragmatic, and causal view of the mind, so that Walter Freeman's judgment is a misunderstanding:

My conception of the holistic nature of the operation of Intentional states would preclude a mechanical reflex conception at the level of Intentionality, whether conscious or unconscious . . . it is a consequence of my views on the mind-body problem that it would be miraculous if the holistic structure of the mental were realized in neurophysiological systems in a way that did not at all reflect this holism. So, though I did not advance theories in neurobiology, my account of the Intentional, if correct, would make it exceedingly unlikely that the reflex arc is the correct model for all or most brain functioning. (1991b, p. 143)

This citation is Searle's direct answer to Freeman and Christine Skarda's attacks against the theory of causal self-referentiality of action and perception. Searle's analysis of Network and Background shows how a holistic and pragmatic model of mind can also causally characterize causality as an essential feature of the most primitive mental states without involving any form of behaviorism.

4. The Intersubjective Dimension of Individual Consciousness and Intentionality: Problems and First Steps Toward a Theoretical Account

Prima facie the concept of Background might produce some problems for Searle's thesis that the brain is causally sufficient to determine mental states (LePore, 1995). Let us compare the following passages:

An Intentional state only determines its conditions of satisfaction – and thus only is the state that it is—given its position in a *Network* of other Intentional states and against a *Background* of practices and preintentional assumptions that are neither themselves Intentional states, nor are they parts of the conditions of satisfaction of Intentional states. To see this, consider the following example. Suppose there was a particular moment at which Jimmy Carter first formed the desire to run for the Presidency of the United States, and suppose further that this Intentional state was realized according to everybody's favorite theories of the ontology of the mental . . . he had a certain neural configuration in a certain part of his brain which realized his desire. . . . Now suppose further that exactly these same type-identical realizations of the mental state occurred in the mind and brain of a Pleisto-

cene man living in a hunter-gatherer society of thousands of years ago. He had a type-identical neural configuration to that which corresponded to Carter's desire. . . . All the same, however type-identical the two realizations might be, the Pleistocene man's mental state could not have been the desire to run for the Presidency of the United States. Why not? Well, to put it as an understatement, the circumstances were not appropriate. (Searle 1983, pp. 19–20)

Even if I am a brain in a vat . . . nonetheless, I do have the Intentional content that I have, and thus I necessarily have the same Background that I would have if I were not a brain in a vat and had that particular Intentional content. *That* I have a certain set of Intentional states and *that* I have a Background do not logically require that I be in fact in certain relations to the world around me, even though I could not, as a matter of empirical fact, have the Background that I do have without a specific biological history and a specific set of social relations to other people and physical relations to natural objects and artifacts. (Ibid., p. 154)

The concept of Background apparently introduces an external component into the determination of intentional contents because even though we had type identity between neurophysiological states in the prehistoric hunter and the person who wants to be President of the United States, the attribution of the desire, "I want to be President of the United States," is implausible because background external circumstances (the "local Background") does not allow the hunter to think that thought.

But this conclusion is not easily reconcilable with the second passage, in which Searle says that since the Background is a mental and nonrepresentational phenomenon, a completely isolated individual (the brain in a vat) could also have it. The "brain in a vat requirement" (also known as "methodological solipsism") foreshadows, according to many philosophers (Putnam, 1996, pp. xiii–xx; Viskovatoff, 2001), an unacceptable return to a Cartesian solipsistic subjectivism, while others (Habermas, 1991; Apel, 1991) express the same point, saying that the entire project of grounding language on mind leads to the same solipsistic and Cartesian conclusion. McDowell expresses the critique as follows:

We have all kinds of things other than the brain for the purpose of representing the world to ourselves: sounds, marks on paper, arrangements of models . . . in short, all kinds of bits of the world outside our brain. . . . It is as if Searle has here forgotten, or perhaps never quite took the measure of, the "direct realism" that he resolutely espouses . . . about perception . . . Surely no sane naturalism can possibly compel us to accept the idea that being in the world, for *us*, is being inside our own heads. The idea has a comical ring; and the indirectness which it imposes on how we have to pic

ture our dealings with the world . . . sits ill with Searle's earlier commendable rejection of intermediaries (1991, pp. 221–222)

On the other side, Francesca Di Lorenzo Ajello argues (against Jürgen Habermas's and Karl O. Apel's critiques of Searle's supposed solipsism) that these objections misunderstand Searle's theory. She reconstructs the continuity of Searle's theory of intentionality with the speech acts theory and the crucial role played in these inquiries by the paradigmatic role assumed by intentional action and the concept of Background in Searle's biological and cultural dimension. Di Lorenzo Ajello's interpretation finds wide positive confirmation in Searle's theories of social ontology and rationality, where Searle analyzes the intersubjective structure of consciousness.

Apel, Habermas, McDowell, and others raise a general problem: the subjective consciousness has an intersubjective dimension (Liotti, 2005). Searle is aware of this problem, as he writes in some passages in *The Rediscovery of the Mind*:

I am convinced that the category of "other people" plays a special role in the *structure* of our conscious experiences, a role unlike that of objects and states of affairs; and I believe that this capacity for assigning a special status to other loci of consciousness is biologically based and is a Background presupposition for all forms of collective intentionality. . . . But I do not yet know how to demonstrate these claims, nor how to analyze the structure of the social element in individual consciousness. (1992, pp. 127–128)

Searle also states the guidelines for the proper scientific and philosophical study of mind: "A fourth and last guideline is that we need to rediscover the social character of the mind" (ibid., p. 248).

These passages make a commitment to the development of a complex philosophical project on these issues that starts with biological naturalism and the theory of intentionality. We can begin with a discussion of the picture of the brain in a vat. As Giulio Tononi and others recently pointed out (Tononi, 2003a; cf. Llinàs, 1996; 2001) phenomena such as dreams show the compatibility of a view of the brain as an autonomous system, not directed from the outer, with the idea that its tendencies are partly molded by interactions with others and the environment (at the ontogenetic level):

The first lesson of dreams is . . . that the brain produces conscious experiences when it is awake and when it sleeps, regardless whether the external world is in contact with us. Consciousness is a property of the brain, of its organization. Of course, the external world has a role. To be what it is, the brain must have had a long evolutionary history, continuous interactions with the external world, and numerous adaptive variations. It needed to become an adult brain continuously changing by virtue of the experience of the world, of other beings similar to us, and of language. But once it be-

came what it is, the brain has whatever is necessary and sufficient to be conscious during night and day (my translation). (Tononi, 2003a, p. 137)

Under this interpretation, the picture of the brain in a vat is not the contemporary version of Descartes' evil genius because it does not assert that mind is entirely independent of its biological and cultural embodiment. Instead, the causal sufficiency requirement embodied in biological naturalism (brains cause minds) just states, under this interpretation, that mental life occurs in individual organisms. This interpretation holds even for biological naturalism and the theory of intentionality as I can argue with different reasons.

First, Searle's use of the phrase bootstrapping effect denotes the view of the brain as an autonomous system constrained by its embodiment.

Second, this interpretation makes sense of Searle's critique of Noam Chomsky's innatism. If the solipsistic interpretation of Searle's thought holds, then Searle would find in Chomsky's innatism a valid ally. But Searle argues instead for a social conception of language together with a holistic view of mind. The speaker is brought up in a community and learns to master concepts in a culture. In this passage, for example, Searle implicitly uses the concepts of Network and Background:

When Chomsky suggests that the concepts expressed by words like "carburetor" and "bureaucrat" must be innately known by every child, and that learning the meanings of the words is just a matter of applying labels to concepts the child already has, you know that something has gone radically wrong. He has a very unrealistic conception of learning. It is as if he supposed that learning the meanings of these words would have to consist in having one's nerve endings stimulated by passing bureaucrats and carburetors, and because there is no way such passing stimuli could ever give us the meanings of these words, it looks like the meanings must be innate.

This argument is called the argument from the "poverty of the stimulus" and it occurs over and over in Chomsky's work. But a more realistic conception is the following: in order to understand, for example, the word "bureaucrat," a child has to be introduced to a culture, a culture that includes governments, bureaus, departments, powers, employment, and a host of other things. A child does not learn a set of discrete concepts, but learns to master a culture, and once that culture is mastered, it is not difficult for him to understand the word "bureaucrat." Similar remarks could be made about "carburetor." This concept only makes sense within the context of some knowledge of internal combustion engines. Once you have the basic understanding of how such engines work it is not hard to understand that a carburetor is a device for mixing air and fuel. (2002h, p. 35)

Finally, further proofs for this interpretation come from Searle's analysis of the dispute between internalism and externalism.

Externalism is a theory concerning the causal circumstances enabling a subject to determine a content. It asks how is it possible that a speaker-agent could have a given content. Searle's theory, on the other hand, is a theory concerning the *constitution* of the content. It asks what determines the content with respect to others, that could be logically possible and it is an example of internalism because it states that the mind of the speaker has sufficient sources to determine contents (Searle, 2004, p. 188).

Regarding the first question (on the causally enabling conditions), the answer could be extremely complex and we could hardly find a general account in causal terms:

Causally speaking, I do not think there is any general answer to this question except to say that our intentional contents are determined by a combination of our life experiences and our innate biological capacities. I have already given a sketch of how an animal's feeling of thirst might be determined by neurobiological processes. If one were to change the example slightly so that I was not just thirsty in general but thirsty for a glass of draught Irish stout, or a 1953 Chateau Lafitte, then the story would become much more complicated. I would have to give an account of how my life experiences have led me to have certain sorts of taste experiences, that I was capable of recalling these in memory and capable of forming desires to repeat these experiences in the future. But if the story has to be more complicated to account for a specific desire, then it would become incredibly complicated if I tried to give an account of how one might have formed an intention with the content that I write the Great American Novel, marry a Republican, or explain intentionality in a single chapter. (Ibid.)

The constitution of intentional states, on the other side, is determined by elements such as psychological mode, direction of fit, Network, and Background. Among these elements, the concept of causal self-referentiality plays a crucial role. On the contrary externalism states that content-determination depends on external causal circumstances and environmental features. Consider Sir Hilary Putnam's "Twin Earth" thought experiment (Putnam, 1975b). Imagine that, somewhere in the universe, a planet exists, named "Twin Earth", which is type-identical with the Earth. The only exception is that the odorless, tasteless, colorless liquid called "water" in both the planets is H_2O on Earth, but XYZ on Twin Earth. So if we look only at the mental content of a person on Earth and of his twin on Twin Earth when they think "I want a glass of water", we will never know that the references of the thoughts are different. To know the difference we should look outside the speakers' heads and see that the respective mental contents are determined by the features of the environment and by the causal interactions between speakers and environment. Also, we should look at the causal interactions between the speakers, starting from the initial baptism of an indexically determined substance (and of substances type-identical with "this") as water.

Searle's idea is that arguments like this make their point using a view of mental states that assimilates concepts and mental descriptions to Frege's *Sinnen*, or to the universals of the Plato-Aristotle tradition. But the problem is that this view of mental states would be incapable of taking into account the particularity of "this thing here." Consequently, according to the externalist, since every semantic act implies a reference to something indexically determined, we have to draw the conclusion that content determination depends on causal relations outside of the speaker's head.

In Putnam's case, Oscar and twin Oscar may have type-identical mental contents, may also be identical down to the last molecule, but still refer to different things (H_2O on Earth and $X\ YZ$ on Twin Earth): so mental content is not sufficient to determine reference. But the validity of the argument depends crucially on the assumption that mind lacks any indexical dimension at all. But Searle's concept of causal self-referentiality allows us to solve this problem with a deeper analysis of the logical structure of mind (Searle, 1983, chap. 8).

Consider the following point:

For the traditional idea that a checklist of features is associated with each word. . . . Putnam substitutes an indexical definition. . . . On our account of the causal self-referentiality of perceptual intentionality, that amounts to saying that water is whatever is identical in structure with the substance causing this very visual experience. But that definition sets a condition that is entirely represented in the contents of the mind (Searle, 2004, p. 183)

Searle also writes, in more detail:

The argument given by the externalists is in every case the same: two speakers could have type-identical contents in the head and yet mean something different. But the answer given to this claim by the internalists is that in all cases where that is so, it is because there is some indexical component in the head that sets a different condition of satisfaction in the two cases, because it sets the condition relative to the head of the speaker in question. . . . Indexicality will enable type-identical thoughts in the head to determine different conditions of satisfaction because the conditions of satisfaction, being indexically determined, are fixed relative to the head in question. Thus in the Twin Earth case the people on both Earth and Twin Earth set conditions of satisfaction relative to themselves: What we call "water" is anything type-identical in structure with the stuff *we* are seeing. But since the "we" in the two cases is different and since the people on Twin Earth are seeing something different from the people on Earth they will have different conditions of satisfaction even though the contents of the head are type-identical. There is nothing in this example to show that meanings are not in the head. (Ibid., pp. 186–187)

So according to Searle, externalism and internalism would be different answers to different questions. But in the quoted passages, in relation to our problem, we have something more: we can see the claim that a subjective component is necessary for the purpose of content determination, even though we are aware that external items are relevant. But external items are relevant for the speaker-agent only insofar as the item acts on the brain or the agent represents it. When the external items are relevant, this happens because the content is indexically determined, but causal self-referentiality, an internal logical structure of mind, plays exactly this role. It indexically determines a given content in relation to the agent in question.

Even from this point of view, therefore, an internalist view of mind that sees mind as autonomous with respect to the external world is compatible with the thesis that the physical and social environment can shape our subjectivity.

5. Collective Intentionality and Institutional Contexts: Intersubjective Structures of Consciousness

So far, I have explored a potential problem with biological naturalism and the theory of intentionality: the apparent tension between the biological and cultural nature of mind in relation to the brain in a vat requirement and to the concept of Background. Searle's subjectivism, far from involving solipsistic results, foreshadows the possibility of asserting the autonomy and the intersubjective structure of mind.

Searle starts his philosophical project of rediscovering the collective aspect of mind in "Collective Intentions and Actions" (2002d) and develops it in *The Construction of Social Reality* (1995b) but in pursuing this project he is also developing some implicit aspects of his speech acts theory (Di Lorenzo Ajello, 2000; 2003). I will reconstruct the project only within the scope of themes and perspectives relevant to this book, while I refer the reader to other works for a deeper grasp of Searle's "social ontology" (cf. Koeppsell and Moss, 2002; Di Lucia, 2003). Our fundamental question is, how can we account for the intersubjective structure of consciousness and intentionality under the presupposition of the subjective nature of mind?

Recall the example mentioned earlier, where the prehistorical hunter, unlike Hilary Clinton, cannot have the desire to be President of the United States, even though they could have type-identical brain states. Searle's theory would analyze this aspect of mind in the following steps:

- (1) The extension of the theory of intentionality to analyze collective actions. Collective intentionality is the biologically primitive and prelinguistic condition of possibility of collective and cooperative behavior.

- (2) The construction of a theory of institutional facts (such as money, governments, and holidays) through the notions of collective intentionality, imposition of status functions, and constitutive rules.
- (3) The thesis that we can think thoughts such as “I want to be President of the United States” only within an institutional context. This context in turn affects individual minds through language and Background abilities.

The first step recognizes that human beings (and non-human animals), “share intentional states such as beliefs, desires and intentions” (Searle 1995b, 23). The basis of this ability is collective intentionality, a “biologically primitive phenomenon that cannot be reduced to or eliminated in favor of something else” (ibid., p. 24). Each individual can act as part of a collective: people can play the violin part in an orchestra, or make an objection during a conference. The reason is that people have intentional states of the form “we intend” from which they derive intentional states of the form “I intend.” For example, I intend to play a violin solo because we intend to play this movement.

Searle argues not only that “we intentions” are irreducible to sets of “I intentions,” but also that collective behavior cannot be reduced to individual behavior:

[I]magine that a group of people are sitting on the grass in various places in a park. Imagine that it suddenly starts to rain and they all get up and run to a common, centrally located shelter. . . . In this case there is no collective behavior; there is just a sequence of individual acts that happen to converge on a common goal. Now imagine a case where a group of people in a park converge on a common point as a piece of collective behavior. Imagine that they are part of an outdoor ballet where the choreography calls for the entire *corps de ballet* to converge on a common point. We can even imagine that the external bodily movements are indistinguishable in the two cases; the people running for shelter make the same types of bodily movements as the ballet dancers. Externally observed the two cases are indistinguishable, but they are clearly different internally . . . part of the difference is that the form of the intentionality in the first case is that each person has an intention that he or she could express without reference to the others. . . . But in the second case, the individual “I intends”s are . . . derivative from the “we intend”s . . . in the first case, even if each person knows that the other people intend to run to the shelter and knows that the other people know that he intends to run to the shelter we still do not have collective behavior. (2002d, p. 92)

Like any intentional phenomenon, collective intentionality requires special Background abilities, although, as Searle himself says, they are not easy to characterize. We do not require only that a person can move his or her fingers on the violin, but, in a wider sense, that they have “the sorts of things that old time phi-

losophers were driving at when they said things like ‘Man is a social animal’ or ‘Man is a political animal’” (ibid., p.103).

Here is Searle’s hypothesis:

Collective intentionality presupposes a background sense of the other as a candidate for cooperative agency, that is, it presupposes a sense of others as more than mere conscious agents, indeed, as actual or potential members of a cooperative activity. (Ibid., p. 104)

An animal not only will recognize other animals like itself as members of a collective engaged in some collective action. The crucial ability is that each agent takes the others equally as other agents and supposes that they will equally take him or her as a potential candidate for collective action. In this sense, mutual recognition is the basis of the sense of “us” as agents in cooperative activities. This is the reason why society cannot be completely analyzed in terms of cooperative behavior in general and linguistic-conversational behavior in particular:

[I] am here suggesting that we cannot explain society in terms of either conversation in particular or collective behavior in general, since each of these presupposes a form of society before they can function at all. The biologically primitive sense of the other person as a candidate for shared intentionality is a necessary condition of all collective behavior and hence of all conversation. (Ibid., p. 105)

So far, we do not have anything special about human beings in our characterization of cooperative action: collective intentionality and Background are features that we share with other species.

To explain what is typical of human collective behavior and mind we have to add something else:

Animals running in a pack can have all the consciousness and collective intentionality they need. They can even have hierarchies and a dominant male; they can cooperate in the hunt, share their food, and even have pair bonding. But they cannot have marriages, property, or money. . . . Because all these create institutional forms of powers, rights, obligations, duties, etc., and it is characteristic of such phenomena that they create reasons for action that are independent of what you or I or anyone else is otherwise inclined to do. (Searle 1995b, p. 70)

We have a continuum going from the basic ontology of molecules and mountains to the systems of institutions and deontic powers (ibid., p. 41) typical of human civilization. Institutions are those kinds of systems enabling an agent to think, for example, “this is my house,” where the thought implies that I bought the house or constructed it, that I paid taxes, that the house is my property, and

that I have rights to it. To draw a comparison, the members of a group of predators can think “this is our prey,” but not “we have rights to it with respect to the members of the other group.”

In Searle’s terminology, the continuum is marked by the rising of “institutional facts” (property, money, marriage, government, taxes), where these facts exist only by virtue of their being represented as such in a collective agreement, so that they are ontologically subjective. But these facts are also epistemically objective: my bank statement and my status of Italian citizen are not a matter of subjective opinion.

What are the conditions of the possibility of these facts? The decisive move is, Searle writes (*ibid.*, pp. 41ff.), the imposition of a status function *Y* on a physical object *X* through collective intentionality in a context *C*. Through this imposition the object can perform a new function independently of its physical features.

We can state this point with a formula: we create institutional facts through constitutive rules of the form *X* counts as *Y* in *C*. The notion of “constitutive rules” comes from *Speech Acts* (Searle, 1969, pp. 33–42) and refers to those rules that create a new kind of activity, where “regulative rules” govern pre-existing activities. So we could say that some dispositions of pieces (*X*) in a chess match (*C*) count as a checkmate (*Y*) is a constitutive rule, while the rule that the player with white pieces moves first is regulative. Constitutive and regulative rules are also different from conventions, such as the convention that the pieces are black and white (Searle 1995b, pp. 27–29, 43–51).

So we have functions that can be performed by objects (such as people or bits of paper) independently of their physical constitution: a hominid could construct weapons with wood and stones, but something performs the function of being a weapon only because its physical structure allows it to do so. On the contrary, nothing in the physics of a piece of paper makes it money and nothing in the physics of a man or a woman makes him or her a president. The physics of these objects does not justify the system of deontic powers deriving from the imposition of status function. In these cases, only the collective agreement that imposes the function makes it possible for the performance of the function itself.

We can now make explicit the internal relation between collective intentionality, imposition of status functions, and constitutive rules, as follows:

[W]hen we say that such and such bits of paper count as money, we genuinely have a constitutive rule, because satisfying the *X* term . . . is not by itself sufficient for being money, nor does the *X* term specify causal features that would be sufficient to enable the stuff to function as money without human agreement. So the application of the constitutive rule introduces the following features: The *Y* term has to assign a new *status* that the object does not already have just in virtue of satisfying the *X* term; and there has to be collective agreement, or at least acceptance, both in the imposition of that status on the stuff referred to by the *X* term and about the function that goes with that status. (*Ibid.*, p. 44)

So human consciousness and intentionality are intersubjectively structured by collective intentionality and by the imposition of status functions according to the formula of constitutive rules “*X counts as Y in context C.*” We also require Background abilities consisting in taking the other as an actual or potential partner in collective action and in knowing that he has an analogous Background stance toward me.

But this account, which starts with our prelinguistic abilities and ends up with the construction of institutional facts, is still incomplete. Since collective intentionality and imposition of function are abilities we share with other social (but noninstitutional) animals, and since only human beings can impose status functions, we need to understand why they have this ability.

Searle’s thesis on this issue is that those thoughts partly constitutive of social reality (such as “this is money,” or “this is my property”) are language-dependent because the shift from the *X* to the *Y* term in the imposition of status function is not grounded in any physical feature of *X*. We just have to think, or represent, the object *X* as performing the function *Y*. This representation is essentially linguistic, since a defining feature of language is that it has “*conventional devices that mean something or express something or represent or symbolize something beyond themselves, in a way that is publicly understandable*” (ibid., pp. 60–61):

The very design of status-functions is such that they both are partly constituted by thoughts and that prelinguistic forms of thought are inadequate to do the job. The reason is that they exist only by way of collective agreement, and there can be no prelinguistic way of formulating the content of the agreement, because there is no prelinguistic natural phenomenon there. The *Y* term creates a status that is additional to the physical features of the *X* term, and that status has to provide reasons for action that are independent of our natural inclinations. The status exists only if people believe it exists, and the reasons function only if people accept them as reasons. Therefore, the agent must have some way to represent the new status. He cannot do it in terms of prelinguistic brute features of the *X* term. (Ibid., p. 69)

But how do we reconcile this thesis with that of the institutional nature of language? A sound or mark on paper (*X*) counts as a promise (*Y*) in a context *C* only by virtue of a collective agreement on the meaning of those sounds and marks and not by virtue of their physical features. We can answer this objection in different ways. We could point out that higher cognitive abilities require prelinguistic mental abilities and social interaction. This strategy, based on the analysis of the structure of intentional states and on the bootstrapping effect resulting from cooperative interaction, underlies Searle’s answer to the objection, since language:

is precisely designed to be a self-identifying category of institutional facts. The child is brought up in a culture where she learns to treat the sounds that

come out of her own and others' mouths as standing for, or meaning something, or representing something. (Ibid., p. 73)

Language, then, is itself an institution, and we could say, on the grounds of the language-dependence of thoughts constitutive of institutional reality, that "[E]ach institution requires linguistic elements of the facts within that very institution" (ibid., p. 60).

The analysis is not circular because the institution of language is not possible without a rich apparatus of prelinguistic intentionality and without the exposition to cooperative behavior (Searle, 2006). Each one of these conditions reacts on the others giving rise to an organism whose subjective consciousness is not only intersubjectively, but also institutionally structured.

6. Searle's Theory of Intentionality in the View of Contemporary Scientific Research: Action, Perception, and the Background Sense of the Other

Searle's theory of intentionality has some crucial features such as the holistic structure of the Network, its embodiment in a Background of preintentional mental skills, and the biologically and logically primitive role of perception and intentional action (1991d, p. 188).

The holistic interdependence between mental states appears to refute the classical "box and arrow" models of action as the result of the mediation of sensory inputs worked out by cognition, in which perception, cognition, and action are serially ordered and mutually independent items. For example, I grasp a cup of coffee because I see it, I desire a coffee, and I believe that the coffee is in the cup.

But, if Searle is right, the holistic structure of mind cannot allow us to see action, perception, and cognition as independent processes. On the contrary, we should see perception as already mediated by possible action schemes and, in humans, by those concepts and categories whose mastering is part of our (local or deep) Background.

As Gestalt psychologists discovered, all perception is structured in *Gestalten* (semantically loaded figures) working as such only against a background. But, as Searle notes, this phenomenon exemplifies the more general holistic structure of consciousness and intentionality, which articulates individual mental states in an aspectual shape determined by, among other things, the position that the state has in the Network and against preintentional Background abilities. We can account for the figure-background structure in perception as a particular case of a general structure of consciousness and of the formal structure of intentionality:

[T]hese features hang together: structuredness, perception as, the aspectual shape of all intentionality, categories, and the aspect of familiarity. Conscious experiences come to us as structured, those structures enable us to perceive things under aspects, but those aspects are constrained by our

mastery of a set of categories, and those categories, being familiar, enable us, in varying degrees, to assimilate our experiences, however novel, to the familiar. (Searle, 1992, p. 136)

All non-pathological forms of consciousness are experienced under the aspect of familiarity. Because all intentionality is aspectual, all conscious intentionality is aspectual; and the possibility of perceiving, that is, the possibility of experiencing under aspects requires a familiarity with the set of categories under which one experiences those aspects. The ability to apply those categories is a Background ability. (Searle, 1995b, p. 133)

Intentionality and consciousness are aspectual because we always represent something under an aspect or another. The structure of perception, including its being subject to categories, results from the embodiment of perception into a Network of other mental states and from our mastery of linguistic categories, which is part of our Background.

I will explore the impact of the holistic view of consciousness and intentionality on the scientific research of mind in more depth in the next chapter, especially with respect to Edelman and Tononi's "dynamic core hypothesis."

Searle's theory is a general model that can provide a framework for some recent scientific contributions.

As Antti Revonsuo observed, conscious experience in human beings presents a particular form of the phenomenon of "binding" (the unity of consciousness), called semantic-conceptual binding:

As we serially scan our surroundings and shift our attention from one location to the next, not only do the elementary features of objects become bound into coherent entities, but instantaneously we are able to categorize the attended object as a kind of meaningful entity. Thus, each separate object activates a coherent network of semantic knowledge in the brain . . . this type of binding, which I call *semantic-conceptual binding*, certainly is an essential ingredient of our phenomenal consciousness, since it allows us to experience the world around us as familiar and meaningful. (1999, p. 179)

Revonsuo also points out that we have cases of semantic dementia, in which this kind of binding breaks down. These cases do not involve perceptual problems, but produce damage to the correct categorical "mapping" of the world, except in quite generic ways.

Semantic-conceptual binding in relation to perceptual phenomena is an extensively studied phenomenon, but the formal structure of intentionality allows us to interpret these phenomena as a particular case of the more general aspectual shape structured and made possible by the set of Background abilities and by the holistic Network.

Network and Background allow us to give a more general account of the dependence of experiences on a linguistically structured form of expectation (Postman, Bruner, and Walk, 1951; Searle, 1983, p. 53; 1992, p. 136). Semantic binding exemplifies a more general “aspect of familiarity” structuring conscious life, and is made possible by Background abilities.

Other results suggest that seeing the world is not just perceiving it under a set of familiar categories, but as a set of possible actions. Action, perception, and cognition are interdependent and not serially ordered and atomistically separated.

The empirical data are the following: the brain motor areas of human beings and apes have a special kind of neurons called mirror neurons. These neurons have this name because they fire when a subject is engaged in some activity (such as grasping) and when the subject looks at another individual performing the same kind of activity. Empirical tests confirm that these neurons codify perceptual scenes as sets of possible actions.

The recent work on mirror neurons made by Giacomo Rizzolatti and Corrado Sinigaglia (2006) shows that the traditional conception of action is mistaken. Action is not the simple result of information processing carried out by perceptual and cognitive systems in the brain independently of the areas governing the motor system. Instead, the brain motor areas are already functioning in purely perceptual situations, for example when a subject looks at a cup of coffee with the intention to grasp it even though the intention is not carried out.

The perceptual situation is already structured as a set of possibilities of actions, so that the traditional thesis that perception, cognitive mediation, and action are independent fails:

These patterns could be persuasive because we had an extremely simplified picture of the motor system. But today things are no longer that way. We know that this system is composed of a mosaic of frontal and parietal areas closely connected with visual, auditory, tactile areas, and endowed with functional properties more complex than we could ever suspect. We discovered, especially, that in some areas there are some neurons that fire not in relation to simple movements, but in relation to goal-directed motor acts . . . and that they selectively respond to forms and dimensions of objects when we are going to interact with them and when we just observe them. These neurons appear to be capable of discriminating sensory information, selecting it on the basis of the possibilities of action that it offers, independently of whether such possibilities are actually realized or not (my translation). (Ibid., p. 2)

These studies on the structure and function of brain motor systems could provide an empirical justification of Searle’s theses about the holistic interdependence of mental states and in particular of action and perception in the field of consciousness. Holism leads Searle to a radically antiCartesian conception of consciousness that has one of its crucial points in the picture of an acting consciousness:

By consciousness I do not mean the passive subjectivity of the Cartesian tradition, but all of the forms of our conscious life—from the famous “four f’s” of fighting, fleeing, feeding, and fornicating to driving cars, writing books, and scratching our itches. All of the processes that we think of as especially mental—whether perception, learning, inference, decision making, problem solving, the emotions, etc.—are in one way or another crucially related to consciousness. (Searle, 1992, p. 227)

But the studies on mirror neurons and on the motor system also give us other indirect tests, especially on the compatibility between subjective (or individual) nature of mind and its intersubjective (or collective) structure as mediated by the Background ability to recognize others as potential partners in collective actions. Searle says, in particular, that biological naturalism does not imply any return to a Cartesian solipsism, but also that biological and preintentional Background abilities structure the conditions of possibility of collective action and intentionality (and consequently, of institutional and cultural facts in general):

I maintain that in a sense we are, in fact, brains in vats . . . But . . . it is important to notice that I do not say that being in a world consists of being inside our own heads. Rather, what I want to say is, we are *identical* to our own bodies, and the part of our body most important for our Intentional relations with the world is inside our skulls. It is not that, so to speak, we crawl inside our own heads, rather in a sense we *are* our own heads. (Searle, 1991g, p. 240)

In addition to the biological capacity to recognize other people as importantly like us . . . it seems to me that the capacity to engage in collective behavior requires something like a preintentional sense of “the other” as an actual or potential agent like oneself in cooperative activities. . . . (Searle, 2002d, p. 103)

The discovery of mirror neurons confirms that individual brains have the biological abilities that enable them to structure a kind of collective intentionality capable of making collective action possible. In the typical experiment it was observed, for example, that some neurons in an ape’s F-5 area were firing when the animal performed the action to grasp some food, and when the same kind of action was performed by the person conducting the experiment in the animal’s visual field.

We can draw two consequences from these data. The first enforces the thesis that brain architecture does not distinguish sharply between mere perception and mere action because mirror neurons respond also to visually encoded action patterns, but also because perceptual scenes are interpreted by the brain as sets of possible actions. The second consequence turns the attention directly to the thesis I am arguing for: individual brains and minds have the biological basis to pro-

duce mutual recognition between subjects, so that autonomy and intersubjectivity are not incompatible features of mind. As Rizzolatti and Sinigaglia write:

Ranging from the most elementary and natural acts, such as grasping food with hand or mouth, to the more sophisticated acts that require some particular abilities such as making a dance movement, playing a piano part, or a Theatre *pièce*, mirror neurons allow our brain to correlate observed movements to its own and recognizing in this way the meaning of these acts. Without such a mechanism, we could have some “pictorial” representation of others’ behavior, but this would never allow us to know what the others are really doing. Of course, as beings endowed with higher cognitive capacities, we could reflect on what was perceived and infer the possible intentions, expectations, or motivations that could account for the acts made by others. But our brain can understand and recognize these acts immediately, without any kind of reasoning, only on the basis of its motor competences (my translation). (Rizzolatti and Sinigaglia, 2006, pp. 3–4)

In a slogan, as the scientists write, the acting brain “is also and above all an *understanding brain*. It is . . . a pragmatic, preconceptual, and prelinguistic understanding, but not less important because many of our so much celebrated cognitive capacities rely on it” (ibid., 2006, p. 3).

The analysis of data thus far has shown that Searle’s theory of intentionality finds some confirmations in current scientific research.

I will explore this issue in greater depth in the next chapter, with a critical assessment of Searle’s philosophy of mind with respect to contemporary scientific research on consciousness and in relation with the philosophical and scientific issue of the self. The goal will be to explore whether Searle’s model can give valid answers to particular issues and whether we can consider it as a valid general framework for scientific and philosophical reflections on mind.

Five

JOHN SEARLE AND CONTEMPORARY NEUROSCIENCE: HOLISM, MENTAL CAUSATION, AND THE ROOTS OF SUBJECTIVITY

The results of the analysis of John R. Searle's philosophy of mind pointed out, among other things, three crucial elements: the holistic unity of consciousness and intentionality, its evolutionary efficacy, and the causal self-referentiality of perception and action. In this chapter, I will assess the theoretical value of these elements in the context of the contemporary scientific and philosophical debate on consciousness and the self. I will first summarize some crucial points.

Searle contributed to the debate on the unity of consciousness with his analysis of the logical relation between subjectivity, qualitativeness, and unity. We cannot have qualitative experiences without an experiencing subject and we cannot have experiences that are not holistically unified with other experiences within a subjective field. "Unity," according to Searle, means that every conscious state is what it is only given its position within a holistically articulated field.

The implication of this thesis is the possible failure of scientific research projects on consciousness that rely on a "building block" approach to consciousness. This approach states that for every qualitative state, we should identify its localization in the brain's anatomy (the so-called neuronal correlate of consciousness [NCC]) and next find the device binding the building blocks (binding device).

Other scientists developed an alternative, holistic approach. Gerald M. Edelman and Giulio Tononi, for example, explain the properties of integration and differentiation of consciousness hypothesizing that the process generating consciousness is at the level of the "dynamic core," a set of complex interactions realized by means of reentrant mechanisms mainly in the thalamocortical system.

The second element, the evolutionary efficacy of consciousness, is related to the problem of mental causation and to the problem of the unity of consciousness. Antonio Damasio hypothesized that an organism constantly engaged in renegotiating its relations with the environment would have a high evolutionary advantage if it developed a device structuring the organism-environment relation around a space of self-referentiality. The device, according to Damasio, is consciousness. Thanks to consciousness, the organism perceives needs and actions as its own and structures its relations with the environment in a richer, powerful, and creative way. A conscious organism would maximize the strategies of action, making them more flexible and creative, because they would be perceived by that particular organism as belonging to it.

I want to make explicit the theoretical perspectives coming from these considerations. The holistic approach to the problem of consciousness adopted by

Edelman and Tononi could be an empirical confirmation of biological naturalism. The argument for this point is that the holistic approach implies a refutation of the reductionism of the atomistic approach. Holism explicitly treats conscious states as processes and adopts a version of the desubstantialization of mind underlying biological naturalism. Like biological naturalism, the holistic approach conceives of the integrated and differentiated field of consciousness as a state of the brain, realized at an ontologically irreducible higher level of its organization.

In addition, Damasio's hypothesis on the causal and evolutionary powers of consciousness introduces a specification to the concept of consciousness as a holistically unified field. The field gives an organism the power of experiencing itself as causing or caused by states of affairs in the world because it contains a minimal sense of the self. I think that this thesis is present *in nuce* in Searle's analysis of the logical structure of intentional states.

Content determination depends on a holistic Network of other intentional states and on a Background of preintentional skills and abilities grounded on the particular biological structure of the organism and in local, cultural contexts.

Furthermore, the primitive biological components of the Network (perceptions, intentions, and memories) have a causally self-referential element: the contents refer to the state as causing (intentions) or caused by the conditions (perceptions, memories) of satisfaction specified by the content.

The biological, embodied nature of consciousness articulates its holistic structure in a way that produces, to use Damasio's terminology, a core-self, the mere "feeling of what happens" to this very organism in the limited dimension of the "here and now." This primitive self maximizes the behavior of the organism, giving flexibility to the actions. As Damasio says, then, understanding the holistic structure of consciousness is not sufficient to understand its nature and function. A scientific theory, instead, has to show how the biological embodiment of consciousness creates that fruitful interaction between holism and the causal, evolutionary power out of which all of the complex and articulated forms of the self-conscious and reflective "I" come from:

[A] theory of consciousness should *not* be just a theory of how the brain creates integrated and unified mental scenes, although the production of integrated and unified mental scenes is an important aspect of consciousness, especially at its highest levels. Those scenes do not exist in a vacuum. I believe they are integrated and unified *because* of the singularity of the organism and *for* the benefit of that single organism. (Damasio, 1999, pp. 18–19)

1. Building Block Approach vs. Unified Field Model: Implications of the Interdependence of Qualitativeness, Subjectivity, and Holistic Unity of Consciousness for Scientific Research

The logical relation between qualitativeness, subjectivity, and unity of conscious experiences implies some consequences for scientific research on consciousness. If we assume the unity of consciousness and its holistic structure as a target of scientific research, we can think that the atomistic research projects can have little success. The holistic structure of consciousness would require the refusal of those approaches grounded on the idea that first we have to find the NCC for each single conscious experience and then construct the unified field out of the building blocks. Here is what Searle writes on the issue:

I want to suggest that if we think of consciousness holistically, then . . . our original combination of subjectivity, qualitativeness, and unity all in one feature, will seem less mysterious. . . . The research implication of this is that we should look for consciousness as a feature of the brain emerging from the activities of large masses of neurons, and which cannot be explained by the activities of individual neurons. (2002c, p. 56)

Consider a research project grounded on Larry Weiskrantz's studies on blindsight. Weiskrantz's patients have a brain damage in the visual cortex that causes blindness in part of the visual field. But if we place an object in the blind part of the visual field and ask the patients to say what kind of object it is (under a binary choice) they will succeed in identifying the object with a greater probability than is attributable to mere chance. They are not aware of their success (they will say something like "I guess it is an O, but I can't see anything"). In some sense, Weiskrantz's patients "see" the objects although they do not have any qualitative experience of sight.

In such cases, the difference between a conscious brain and a simple information-processing device is well exemplified. Blindsight patients have partial information of the environment from their blind fields but this information is not phenomenally realized. At this point, we can ask what a normal brain has that the impaired one lacks; we can look for the NCC. The next steps would include testing the correlation to see if it is causal and working out a theory that systematically accounts for the phenomenon.

According to Searle, famous neuroscientists such as Francis Crick, Christof Koch, Semir Zeki, and others are working within this atomistic paradigm (Crick, 1994; Zeki, 2003; Koch, 2004). Searle too, in *The Mystery of Consciousness*, regarded the implications of blindsight as the most promising paradigm for the study of consciousness (1997, p. 199–200). But he expresses some doubts about the atomistic approach, drawing a comparison between this model and what he calls the "holistic" or "unified field" approach:

- (1) The building block theory: The conscious field is made up of small components that combine to form the field. To find the causal NCC for any component is to find an element that is causally necessary and sufficient for that conscious experience. Hence, to find even one is, in an important sense, to crack the problem of consciousness.
- (2) The unified field approach . . . Conscious experience comes in unified fields. In order to have a visual experience, a subject has to be conscious already and the experience is a modification of the field. Neither blindsight, binocular rivalry, nor normal vision can give us a genuine causal NCC because already conscious subjects can have these experiences (2002c, pp. 53–54)

Searle doubts that these research projects can answer a more general question: what is it that makes the brain capable of causing consciousness? But the revival of the idea that consciousness is a dynamic process, together with the low explanatory value exhibited by these “atomistic” research projects, such as those grounded on Crick’s and Koch’s 40 Hz hypothesis (Crick, 1994; Crick, Koch, 1997) forced some neurobiologists to shift their research projects from local realization to systemic interaction. Among these, Antti Revonsuo (1999) and Jean Pierre Changeux (2005), for example, explicitly mention Searle and his thesis that a holistic view of consciousness could be the key for a new scientific approach on the problem. Rodolfo Llinàs works out a holistic theory of qualia (Llinàs, 2001, chap. 10; Llinàs and Parè, 1996) and edited an anthology with Patricia Churchland on the holistic view of the mind-brain (Llinàs and Churchland 1966).

If we have to conceive of consciousness as a dynamic process within which single experiences do not act as generators of the fields, but as modulators of it (Llinàs, 2001), the reductionism implicit in the atomistic approach (one NCC = one quale) loses its appeal. We cannot understand by virtue of what kind of local or material property an NCC realizes a given quale instead of another. Also, the hypothesis that an otherwise unconscious living brain can have the power to produce a single experience and nothing else, because the NCC fires, appears to be extremely implausible.

Finally, the realization basis of a conscious experience is not by itself sufficient to produce that experience unless it satisfies some conditions in the total state of the system. This means that the correct functioning of a given brain area associated with the production of a determined conscious experience is not by itself sufficient, though probably necessary, to produce the experience. As we can see from studies of (among others) split-brain patients, if the relevant neuronal group is functionally or anatomically disconnected from the rest of the brain system, or from the thalamocortical system, the relevant experience cannot be produced, notwithstanding the perfect functioning of the area atomistically considered (Edelman and Tononi, 2000; Searle, 2002c, pp. 54–58).

The hypothesis that consciousness is composed of atoms or isolated experiential building blocks is, then, implausible. But we still have to see whether Searle's holistic hypotheses can provide a useful framework to understand the shift from atomistic to holistic approaches in the scientific study of consciousness.

2. Edelman, Tononi, and the Dynamic Core Hypothesis

In *A Universe of Consciousness* (2000), Edelman and Tononi integrate Edelman's neuronal Darwinism theory with an explicit critique of the building block model of consciousness, which they compare with Edward Titchener's old atomistic approach to the physiology of sensation (see also Tononi, 2003a; 2003b) as the search of the single atoms of experience. They instead explain the complex network of integrated and differentiated perceptions and thoughts in the field of consciousness through the dynamic core hypothesis.

According to this hypothesis, integration would be realized through reentrant mechanisms mutually linking neuronal maps distributed in the thalamocortical system. These mechanisms would define a set of synchronized maps with different functions within a single functional cluster. This set of maps and mechanisms, which is changing in composition and does not have a defined spatial location, is the dynamic core.

In the following passage, Edelman states his thesis of what he names "primary consciousness" (the set of perceptions, images and sensations lacking self-consciousness and a socially and linguistically defined sense of the self):

[T]he brain carries out a process of conceptual "self-categorization." Self-categories are built by matching past perceptual categories with signals from value systems, a process carried out by cortical systems capable of conceptual functions. This value-category system then interacts via reentrant connections with brain areas carrying out ongoing perceptual categorizations of world events and signals. Perceptual (phenomenal) experience arises from the correlation by a conceptual memory of a set of ongoing perceptual categorizations. Primary consciousness is a kind of "remembered present." (1992, pp. 119–120)

Edelman's idea is that a fundamental feature of primary consciousness is the power to construct a coherent scene. This scene is characterized by integration and differentiation: the brain integrates incoming signals with categories and concepts constructed in the past and restructures the synaptic connections between relevant neuronal populations through reentrant connections. This allows different brain areas to exchange and integrate signals. The present (the construction of the conscious awareness here and now) is always remembered, since the here and now is constructed through categorical and conceptual sources that ontogenetic and phylogenetic history made available. This is not a one way process:

categories are not hard-wired patterns; instead, memory re-categorizes its categories and concepts in the interactions with ongoing categorizations.

Even in the construction of the vertical or synchronic unity, we have a temporal or diachronic dimension. Edelman supports this hypothesis with the details of how different memory systems are realized in neuroanatomy. As Searle writes, Edelman's model and the view of memory as an active system are one of the main contributions to the contemporary debate on consciousness:

Memory for Edelman is not just a passive process of storing but an active process of recategorizing on the basis of previous categorizations. . . . suppose an animal has acquired a perceptual category of cats. It acquires this category through the experience of seeing a cat and organizing its experience by way of the reentrant map. Then the next time it sees a cat, and thus has a similar perceptual input, it recategorizes the input by enhancing the previously established categorization. . . . It does not just *recall a stereotype* but continually *reinvents* the category of cats. This conception of memory seems to me one of the most powerful features of the book because it provides an alternative to the traditional idea of memory as a storehouse of knowledge and experience, and of remembering as a process of retrieval from the storehouse. (1997, p. 44)

The dynamic core hypothesis generalizes Edelman's theory of perceptual categorization. As Edelman describes it in his model of primary consciousness, perception is a process dependent on many other processes and cognitive abilities whose neural bases are at one time distributed in the brain and integrated by reentrant connections. Likewise, consciousness is a holistic process dependent on many integrated and differentiated, neuroanatomically distributed and functionally integrated processes. Here is a summary of Edelman's approach:

[A]ssuming that certain local properties of neurons may sooner or later hold the key to the mystery of consciousness is entirely unsatisfactory. How could having a specific location in the brain, firing in a particular mode or at a particular frequency, being connected to certain other neurons, or expressing a particular biochemical compound or gene endow a neuron with the remarkable property of giving rise to conscious experience? The logical and philosophical problems of hypostatization associated with such assumption are all too obvious, as philosophers and scientists have noted many times.

Consciousness is neither a thing nor a simple property. Instead, our approach has been to focus on the fundamental properties of conscious experience—such properties as integration and differentiation—and explain them in terms of neural processes. (Edelman and Tononi, 2000, p. 143)

The involved neural processes are in turn integrated, differentiated, and distributed. We can see this point with a close examination of syndromes resulting from brain damage. It can also be proved by the normal functioning of the brain. So for example, in the case of split brain patients, the functional disconnection of cognitive abilities is not explained by some damage to visual, motor or linguistic areas (whose correspondent competences are usually impaired), but by the elimination of the corpus callosum, which cuts the functional and neuroanatomical connections between different areas.

Also, although the brain does not have some hard-wired modular architecture, the brain does not work as a homogeneous and undifferentiated whole. The undifferentiated functioning of the brain, instead, is typical of situations associated with absence of consciousness, such as global epilepsy and dreamless sleep.

Integration and differentiation are the parameters articulating the holistic structure of conscious experiences.

[C]onsciousness is unitary or integrated, at least in healthy individuals. When I consider my conscious state at the time of this writing, it appears to be all of a piece. While I am paying attention to the act of writing, I am aware of a ray of sunlight, of a humming sound across the street, of a small discomfort in my legs at the edge of the chair, and even of a “fringe”, as [William] James called it, that is, of objects and events barely sensed. It is usually not entirely possible to reduce this integrated scene to just one thing. . . . Yet this unitary scene will change and differentiate according to outside stimuli or inner thoughts to yet another scene. The number of such differentiated scenes seems endless, yet each is unitary.” (Edelman, 2004, pp. 7–8)

Edelman and Tononi work out two explanatory hypotheses about the relations between consciousness and the brain to explain the empirical data:

- (1) *A group of neurons can contribute directly to conscious experience only if it is part of a distributed functional cluster that, through reentrant interactions in the thalamocortical system, achieves high integration in hundreds of milliseconds.*
- (2) *To sustain conscious experience, it is essential that this functional cluster be highly differentiated, indicating high values of complexity.*

We call such a cluster of neuronal groups that are strongly interacting among themselves and that have distinct functional borders with the rest of the brain at the time scale of fractions of a second a “dynamic core” to emphasize both its integration and its constantly changing composition. A dynamic core is therefore a process, not a thing or a place, and it is defined in terms of neural interactions, rather than in terms of specific neural location, connectivity, or activity. Although a dynamic core will have a spatial extension, it is, in general, spatially distributed, as well as changing in compo-

sition, and thus cannot be localized to a single place in the brain. Furthermore, even if a functional cluster with such properties is identified, we predict that it will be associated with conscious experience only if reentrant interactions within it are sufficiently differentiated, as indicated by its complexity. (Edelman and Tononi, 2000, p. 144)

The conception of consciousness as a holistically structured process defines empirically testable working hypotheses and, according to the scientists, it is capable of solving the problems of the building block approach. The *explanandum* of the theory is not just a set of objective, third-personal functional roles. Instead, the *explananda* are conscious states as such—subjective and qualitative processes. As Edelman writes, the embodied, biological nature of consciousness is closely related to the essential features of consciousness itself, namely unity, subjectivity, and qualitateness:

As Sherrington recognized, the private nature of conscious events is closely coupled with their unity or their integration. Saying that a conscious state is unified and integrated simply means that the whole experienced conscious state is always more than the sum of its parts. . . .

Another way to say this is that a particular conscious state consists of a tightly interwoven set of relationships that cannot be fully broken down into independent components. (Ibid., p. 24)

While Tononi insists mainly on the abstract, functional definition of the theory through systems of equations based on the theory of complex systems (Tononi 2003a; 2003b), Edelman focuses on the concrete biological instantiation of conscious states (the embodiment of consciousness in a brain and in an organism) and on their subjective nature.

We can see these theoretical proposals as indirect confirmations of Searle's views on the structure, functioning, and ways of realization of conscious processes in the natural world. Like Searle, Edelman and Tononi look for the biological basis of consciousness conceiving of conscious phenomena as subjective, embodied processes that, as such, are irreducible to their basis. The biological basis of consciousness is itself a process that does not have a single place as its "seat," so that the subjectivity of consciousness does not imply a Daniel C. Dennett or René Descartes' style Cartesian Theater.

Also, like Searle, Edelman and Tononi see a close relation between the non-reductive embodiment of consciousness and the features of subjectivity, qualitateness, and unity. Consciousness is subjective, qualitative, and unified *because of* its embodiment in the biological make-up of a single organism.

In addition, the holism of consciousness, the interdependence of conscious and mental phenomena in an integrated and differentiated conscious field, which is supported by Edelman and Tononi with empirical tests and clinical reports,

finds in Searle's model a general framework worked out in terms of philosophical, logical analysis.

The insistence on the scientific explainability of consciousness as such and its irreducibility to single brain places defines the conjugability of naturalism and antireductionism, of ontological subjectivity, and epistemic objectivity.

Edelman's and Tononi's indirect confirmations are noteworthy not only for the relation between holism, biological nature of consciousness, and subjectivity, but also for the relation between the concept of dynamic core and of the self grounded on the hypothesis that the most sophisticated cognitive abilities have their roots in primitive powers mediated by the structures in the brainstem:

[S]tructures in the brain stem map the state of the body and its relation to both the inside and outside environment on the basis of multimodal signals that include proprioceptive, kinesthetic, somatosensory, and autonomic components. We may, indeed, call these components the dimensions of the protoself. These components are the bodily functions of which we are usually only dimly aware, but that influence almost every aspect of our being. . . . Since, memory is recategorial and there is a constant play in time between value-category systems and ongoing perceptual categorizations, this early, body-based consciousness may provide the initial dominant axes of the *N*-dimensional neural reference space, out of which all subsequent memories based on signals from the world . . . are elaborated. (Edelman and Tononi, 2000, p. 174)

As we will see, this thesis, although anticipated by Edelman (1989; 1992) with his analyses of the biological basis of the distinction between self and non-self, has been fully developed by Damasio (1999). For him, the proto-self could be the root of the most sophisticated activities of a self-conscious self. He also maintains that the development of the self is crucial from an evolutionary point of view. From my point of view, Searle's philosophy of mind can give a framework to these hypotheses and achievements when we conjugate the interdependence between qualitateness, subjectivity, and holistic structure with causal self-referentiality, the Network, and the arguments for the evolutionary role of consciousness.

But before I explore Damasio's contribution to the issue of the relations between consciousness and the self, I need to sketch the philosophical issues involved to appreciate the difference between Searle's perspective and the orthodox eliminativistic mainstream.

3. From Consciousness to the Self: Some Possible Post-Humean Eliminative Trends

From the phenomenological point of view, we can say that a sense of the self is typically present in the field of consciousness (Searle, 2004, pp. 144–145). For

example, a pain is not free in the air because it is *my* pain, the pain *I* have relative to a part of my body. Also, theoretical and practical problems do not reflect on themselves. Instead, each one of us is engaged in reasoning with the goal of problem solving or decision-making.

Reasoning is “an activity that actual selves engage in . . .” (Searle, 2001, p. 135). Apparently, therefore, the “subject of our attribution of psychological properties” (Searle, 2004, pp. 280–281) exists, a subject that feels something like being itself and without which mental states could not exist once we recognize their subjective ontology.

But what is the reference of this phenomenological sense of the self? According to Descartes, the problem is easy to solve: the “I” is the thinking substance, which is simple and indivisible as a matter of definition. The thinking substance is the immortal soul, ontologically independent of the body. To account for the interaction between soul and body, and for the sense of intimacy that anyone has with his or her body, we could think of the soul as having a seat in the one brain organ that lacks a twin: the pineal gland. In his religiously inspired thought, since the gland and the soul have the property of unity, though predicated in different senses, Descartes postulated that the soul must be seated in the pineal gland (1989, pt. 1, art. 30–35). His reference for “I” then is the soul of the Platonistic-Christian tradition: a non-physical thing which is immortal, immaterial, continuous in time, and ontologically independent of nature and body (for the history of the problem in modern philosophy see Di Francesco, 1998, chap. 1–4).

At the dawn of modern science, a solution grounded on the traditional metaphysical concept of substance and on the religious concept of the soul was, at best, suspicious for many philosophers. If the soul is not a reliable reference for the self, maybe we could be forced to forget the concept of the self altogether:

Since the dawn of modern science in the seventeenth century, there has been nearly unanimous agreement that the self, whatever it is, would be invisible under a microscope, and invisible to introspection, too. For some, this has suggested that the self was a nonphysical soul, a ghost in the machine. For others, it has suggested that the self was nothing at all, a figment of metaphysically fevered imaginations. And for still others, it has suggested only that a self is in one way or another, a sort of abstraction, something whose existence was not in the slightest impugned by its invisibility. After all, one might say, a center of gravity is just as invisible—and just as real. Is that real enough? (Dennett, 1991, pp. 412–413)

We could say that in the modern post-Cartesian era, the self, mind, and consciousness are desubstantialized and that with David Hume, we could agree that the desubstantialization of the self led to a radical dissolution of its existence. Ideas and impressions, Hume says, are “different, and distinguishable, and separable from each other, and may be separately consider’d, and may exist separately, and have no need of any thing to support their existence” (1984, p. 300).

We are “nothing but a bundle or collection of different perceptions, which succeed each other with an inconceivable rapidity, and are in a perpetual flux and movement” (ibid.).

Hume writes that when he goes inside of what he calls “himself” he can never find himself as such, but always finds only particular perceptions. But here he is not formulating some kind of empirical prediction or an autobiographical tale. Instead, he is saying that nothing could satisfy Descartes’ constraints on the notion of the self. Even if we had a particular experience during all of our lives, still that would be just another experience like any other (Searle 2004, p. 292).

Searle summarizes what he regards as the contemporary conclusions of Hume’s critique:

In summary the “self” is entirely reducible to simpler elements. It consists of conscious feelings, including memories and a sense of “me-ness.” . . . These are caused by and are realized in a continuously existing physical system, my embodied brain. On the neo-Humean view, *in addition to all of that there simply is no such thing as the self*. End of the story about the self. (2001, p. 79)

But the most recent results of the debate also deny the importance of the constraints of embodiment and unity. Dennett, for example, is one of the main supporters of this neo-Humean line of thought. He thinks that the “pandemonium” of agencies and contents fighting for control of brain and behavior creates a “virtual captain of the crew,” a useful fiction that serves to maximize the stability of an organism’s behavior and of others’ expectations. Just as centers of gravity are not physical points but useful devices for prediction, so too do brains “spin” (Dennett 1991, 413) a center of narrative gravity, inheriting this power from the primitive biological instinct to protect and extend the bodily boundaries (ibid., 415–417). But unlike their phylogenetic ancestors, human beings live in an environment whose scene is dominated by culture and language:

Our human environment contains not just food and shelter, enemies to fight or flee, and conspecifics with whom to mate, but words, words, words. These words are potent elements of our environment that we readily incorporate, ingesting and extruding them, weaving them like spiderwebs into self-protective strings of *narratives*. . . .

Our fundamental tactic of self-protection, self-control, and self-definition is not spinning webs or building dams, but telling stories, and more particularly concocting and controlling the story we tell others—and ourselves—about who we are. . . . Our tales are spun, but for the most part we don’t spin them; they spin us. Our human consciousness, and our narrative selfhood, is their product, not their source. (Ibid., pp. 417–418)

Consider two points in Dennett's passage. First, the passive character of the virtual self: the tales constructed by our brains spin us. "We" are their product and not the producers. The virtual captain is not an executive agent, but just a representative character resembling Elizabeth II instead of George W. Bush or Tony Blair (*ibid.*, p. 416).

Second, just as consciousness is not a real and unified stream since it could not have its seat in any single place in the brain, so too is the self a virtual, narrative product of the fight for "brain celebrity." The self has no causal power, no "government power" on behavior. To support the reality of the selves, we would have to believe, again, in the Myth of the Cartesian Theater. As proof of this thesis, Dennett describes cases of identity dissociation, split-brain patients, and many others, as well as a case in which, surprisingly, two bodies appear to be governed by a single self.

Consider for example, the case of the Chaplins. These twin sisters described by Dennett seem to act as one:

[T]hey collaborate on the speaking of a single speech act, for instance, finishing each other's sentences with ease or speaking in unison, with one just a split-second behind. . . . Some who have dealt with them suggest that the natural and effective tactic that suggested itself was to consider *them* more of a *her*. (*Ibid.*, p. 422)

Dennett's conclusion is the elimination of the self. For Dennett, if we want to account for these cases, we have to abandon the traditional picture of mind as a ghost in the machine, and to allow for the existence of quasi-selves and transitional selves. This move would be allowed by Dennett's Multiple Drafts Model, and it could serve to explain identity dysfunctions and the case of the Chaplin twins, whose behavior is due presumably to their history:

[T]hese twins have seen, heard, touched, smelled, and thought about very much the same events throughout their lives, and started, no doubt, with brains quite similarly disposed to react to these stimuli, it might not take enormous channels of communication to keep them homing in a sort of loose harmony. (*Ibid.*)

The elimination of the criterion of embodiment as relevant for the definition of the nature of the self gave rise to a new model (Clark and Chalmers, 2002; for a critical assessment of the model see Di Francesco, 2004), known as "the extended mind," put forward as an interpretation of the concept of mind underlying research in post-classic cognitive sciences. As Andy Clark and David J. Chalmers say, human agents tend to "lean heavily on environmental supports" (2002, p. 644) to carry out cognitive processes. To explain this tendency, we have to invoke an extended notion of mind: "external supports" (paper, pen, books, computers) are mental as much as "internal supports" (neurophysiological

and cognitive individual states) are. The environment is not a passive set of non-mental things; instead it is mental because without it and its features we could not carry out our cognitive activities. Clark and Chalmers give many examples to clarify and argue this thesis. For example, they describe the results of a psychological research on subjects playing the videogame Tetris, where the player has to fit geometrical bi-dimensional forms falling down from the top of the screen at increasing speed with the structure emerging from the bottom. The game requires the ability to mentally rotate the forms to fit the structure, but the player can also push a button to physically rotate the forms. The research shows that the mental rotation requires about 1000 milliseconds for an angle of 90°, whereas the physical rotation requires about 300 milliseconds, including the time to press the button.

External structures, therefore, make the processes enormously faster and more powerful: our general tendency is not only to “store” data and information in these external supports, but also to use these sources to carry out processes (pocket calculators, computers, electronic simulations):

Within the lifetime of an organism, too, individual learning may have molded the brain in ways that rely on cognitive extensions that surrounded us as we learned. Language is again a central example here, as are the various physical and computational artifacts that are routinely used as cognitive extensions by children in schools and by trainees in numerous professions. In such cases the brain develops in a way that complements the external structures, and learns to play its role within a unified, densely coupled system. Once we recognize the crucial role of the environment in constraining the evolution and development of cognition, we see that extended cognition is a core cognitive process, not an add-on extra. (*Ibid.*, p. 646)

On this account, mind is not confined within the boundaries of the body. Instead, it results from the combination of internal and external sources where the connections between them have to respect some criteria (*ibid.*, p. 649).

- (1) First, the organism makes a constant use of the external supports.
- (2) The information stored in the supports has to be directly available without difficulties.
- (3) The information is automatically endorsed to carry out the process.
- (4) The information was consciously assumed in the past by the subject although, the authors write, we could get rid of this criterion and define the extended mind only in terms of unconscious mental processes.

On this perspective, the mind is the system of organism and external supports, so that it makes sense to talk about beliefs, thoughts, and desires as spread in the world, but it also makes sense to talk about an extended self:

The moral is that when it comes to belief, there is nothing sacred about skull and skin. What makes some information count as a belief is the role it plays, and there is no reason why the relevant role cannot be played only from inside the body. (Ibid., p. 648)

What, finally, of the self? Does the extended mind imply an extended self? It seems so. Most of us already accept that the self outstrips the boundaries of consciousness; my dispositional beliefs, for example, constitute in some deep sense part of who I am. If so, then these boundaries may also fall beyond the skin. (Ibid., p. 650)

To delve into a deep and critical assessment of the issue of whether the crisis of the substantialist, Cartesian view of mind necessarily implies the virtuality of the self and its elimination (Dennett) is beyond the scope of this work. Neither can I examine in depth the dissolution of the difference between a subject using external devices in its interaction with the environment and a single subject made out of these interactions. To accomplish such an analysis, I would need, among other things, to examine and assess the crucial issues of personal identity and of the status of the person between individual construction of life plans and social construction of individuality (Bodei, 2003; Sparti, 1996; 2001).

But we can notice that the price paid for the perspectives of Dennett, Clark, and Chalmers is quite high with respect to the ordinary conception of ourselves. As Michele Di Francesco writes:

[T]he mind that we expand . . . is not . . . what we could call our *personal mind*. This has peculiar features: it seems to involve the reference to a *subjective ontology* . . . or better . . . the boundaries of a subjective space. To speak about it we have to use an intentional language, which can give voice to the presence of a point of view, of a singular perspective, which is essentially situated and *individual*. It also shows a peculiar unity (and continuity). Its conscious contents are part of a single phenomenological field; they are immediately and without doubt given to the subject, they have a variety of mutual conceptual and motivational relationships. . . .

The virtual nature of the I—to use Dennett’s terminology . . . is not about—the emergence from the sub-personal level, instead it marks such a distance from the ordinary notion that it makes preferable to talk about an *elimination* of the subject instead of a redescription of it. (2004, pp. 122–123)

[W]hat we leave out . . . is the person, namely the notion of an experiencing and acting subject, reasonably unified and integrated, to which mentalistic predicates (of the folk psychology) are attributed. These are used to explain both “what it is like” . . . to be a certain subject and to identify the *reasons* for that subclass of physical happenings that we call “behaviors” because

they are caused by mental states, namely they are generated by a subject. . . . We can talk this way: if cognitive mind has something to do with personal mind at least insofar as it is necessary to explain the emergence of the realm of personal experience and intentional action . . . we have to require something more than a mere causal connection among information processings. We need that mental states be a part of the emergence of a subject whose mental organization of which has the features of unity, continuity, self-accessibility that we mentioned. (Ibid., p. 128)

The epistemic issues involved in the view of mind correctly criticized by Dennett (such as the issue about the dogma of incorrigibility) notwithstanding, the implications of his positions are controversial under the ontological aspect. Dennett implies that, to be “real,” every higher-level phenomenon must be reducible to or identifiable with some particular “place” or “thing” at the lower level. This is his position when he says, for example:

There is no single, definitive “stream of consciousness,” because there is no Central Headquarters, no Cartesian Theater where “it all comes together” for the perusal of a Central Meaner. (Dennett, 1991, 253)

But the implication is subject to counterexamples. Mass-energy points, fields of force, and the force of gravity are not reducible to any place at the lower level, but their reality is not questioned. The same holds for cognitive processes. Language and perception, for example, are based on the correct functioning of the brain, but no single seat of language or perception exists in the brain, since these capacities depend on systemic interaction (Freeman, 2000). Regardless, language and perception are real.

The extended mind model refutes the solipsistic view of mind and correctly points out that the very structure of mind is shaped by the interactions with the physical and social environment. All the same, it has the unacceptable consequence of confusing a subject using external sources to drive its actions with a subject constituted by causal interactions, so that it cannot explain the actual differences between information conveyed by internal mental states and information stored in external sources.

Following Clark and Chalmers’s thought experiment (Clark, Chalmers, 2002, pp. 647–649), suppose that Otto is a patient with Alzheimer’s Disease and serious impairment to his memory, forced to drive his actions using a paper notebook in which he stores all the relevant information. According to Chalmers and Clark, the notebook is now part of Otto’s mind because the causal relations between Otto and the notebook are functionally equivalent to those between normal people and their brain areas devoted to memory.

But suppose that Otto does not recognize the information stored as written by him. Then the notebook would no longer be part of his mind. Notwithstanding the functional equivalence between normal people with their internal states and

Otto with his notebook, we could hardly imagine a case in which a normal person cannot recognize a memory as his or her own. This is an asymmetry with respect to the equivalence stipulated by the extended mind model, and seems to refute the supposed “widespreadness” of our thoughts all around the world, although we can still accept that social and physical interaction is an essential source to shape individual minds (Di Francesco, 2004, pp. 128ff.).

Eliminative positions are apparently made possible by the crisis of the Cartesian picture of mind as a substantial unity separated from the rest of nature. Once that we argue for the process-like nature of mental states and the relevance of the environment for the structure of mind, this point appears sufficient to refute the existence of a connection between mind and individuality that would constitute us as selves.

I argue that this conclusion is not necessary: a connection between embodiment, mental causation, and holism that makes us real selves exists.

Based on my analyses of Searle’s position, I can begin to articulate the reasons for a realist view of the self that, on the ground of the constraint of the embodiment, goes in the same direction as that put forward by Edelman and Damasio.

4. Consciousness as a Space of Self-Referentiality: Damasio’s Theory and the Role of the Proto-Self in Organism-Environment Transactions

The conceptual category of the proto-self in neuroscience refers to the construction in the brain of a biological model of the organism interacting with the world. The proto-self is interpreted as the origin of the most sophisticated cognitive abilities characterizing a “second-order consciousness” including a self-conscious, linguistic, and socially structured self.

The neurobiological theory worked out by Antonio Damasio (Damasio, 1994; 1999; Parvizi and Damasio, 2001; Damasio and Damasio, 1996) extends its scope through the identification of the neural mechanisms underlying its construction within the emergence of core consciousness (which is Damasio’s word for Edelman’s primary consciousness). Also, Damasio’s evolutionary framework allows the neurobiologist to hypothesize a close relation between the embodiment of mind, the development of the self, and the evolutionary role of consciousness. Consciousness, because of its embodiment, would be the wellspring of creativity.

This characteristically Darwinian approach is not by itself distinctive of Damasio’s position with respect to Edelman’s, although recently Edelman accepted a form of epiphenomenalism on the basis of Jaegwon Kim’s causal exclusion argument (Edelman, 2004, chap. 7). But Edelman’s new position hardly coheres with an evolutionary theoretical framework, as Edelman recognized in *The Remembered Present*, when he stated the following condition of adequacy for the theory: “In line with the evolutionary assumption, the theory should consider as causally meaningful conscious and phenomenal states too” (1989, p. 53).

On the contrary, in Damasio's theory, the evolutionary framework coheres with embodiment and causal power of consciousness as constraints on the analysis of the structure of the unified field and gives a basis for the construction of the self.

Damasio's theory allows us to identify some further confirmation of Searle's philosophy of mind in contemporary neuroscience. Also, his neurobiology can give us a key to further test the possible conjunction of naturalism and non-reductionism into a new version of the desubstantialization of mind. I will analyze here some crucial points.

The first is the attribution of a direct causal role to conscious subjectivity, consisting of its ability to creatively manage the organism-environment relationships. In addition, we have a close relationship among the biological conception of consciousness, its causal efficacy, and the emergence of the core-self from the primitive forms of action and perception most linked to the needs of a biological organism.

The issue of the self is the core of Damasio's theory, which divides the problem of consciousness into two parts (the problem of the self and the "hard problem" of qualia) and applies the solution to the neurobiological problem of the self as a meaningful part of the solution to the hard problem:

In the very least . . . the neurobiology of consciousness faces two problems: the problem of how the movie-in-the-brain is generated, and the problem of how the brain also generates the sense that there is an owner and observer for that movie. The two problems are so intimately related that the latter is nested within the former. In effect, the second problem is that of generating the *appearance* of an owner and observer for the movie *within the movie*; and the physiological mechanisms behind the second problem have an influence on the mechanisms behind the first. (Damasio, 1999, p. 11)

The picture of consciousness as a unified field still plays a crucial role in Damasio's theory, but the crucial problem is how a sense of the self is generated in that field under the assumption of the crisis of views hypostatizing mind, consciousness, and the self:

There is no homunculus, either metaphysical or in the brain, sitting in the Cartesian theater as an audience of one and waiting for objects to step into the light. . . . solving the second problem of consciousness consists in discovering the biological underpinnings for the curious ability we human beings have of constructing, not just the mental patterns of an object—the images of persons, places, melodies and of their relationships, in short, the temporally and spatially integrated mental images of something-to-be-known—but also the mental patterns which convey, automatically and naturally, the sense of a self in the act of knowing. Consciousness, as we commonly think of it, from its

basic levels to its most complex, is the unified mental pattern that brings together the object and the self. (Ibid.)

We have different ways to understand why the problem of the self is crucial, and one of the more strategically powerful is to use “the scalpel of neurology.” Damasio asks us first to consider a case of absence crisis followed by epileptic automatism:

Thirty-two years ago, a man sat across from me in a strange, entirely circular, gray-painted examining room. The afternoon sun was shining on us through a skylight as we talked quietly. Suddenly the man stopped, and his face lost animation; his mouth froze, still open, and his eyes became vacuously fixed on some point on the wall behind me. For a few seconds he remained motionless. I spoke his name but there was no reply. Then he began to move a little, he smacked his lips, his eyes shifted to the table between us, he seemed to see a cup of coffee and a small metal vase of flowers; he must have, because he picked up the cup and drank from it. I spoke to him but again and again he did not reply. He touched the vase. I asked him what was going on, and he did not reply, his face had no expression. He did not look at me. . . . Now he turned around and walked slowly to the door. I got up and called him again. He stopped, he looked at me, and some expression returned to his face—he looked perplexed. I called him again, and he said, “What?” (Ibid., pp. 5–6)

The scene is surprising in many respects. The impairment of consciousness is caused by epilepsy in a person who continues to show apparently goal-directed motor and behavioral sequences (such as drinking, touching, and approaching the door). The patient did not lapse into a coma or have convulsions, but, in some sense, “he wasn’t there:”

During the period of impaired consciousness, the man’s wakefulness, his basic ability to attend to objects, and his capacity to navigate in space had been preserved. The essence of his mental process was probably retained, as far as the objects in his surroundings were concerned, but his sense of self and knowing had been suspended. The shaping of my notion of consciousness probably began that day, without my noticing it, and the idea that a sense of self was an indispensable part of the conscious mind only gained strength as I saw comparable cases. (Ibid., p. 7)

The idea is that the existence of the self is, in some sense, the essence of consciousness. The scalpel of neurology in this case helps us to see the distinction between two kinds of consciousness (core and extended) with two different kinds of self and with a well-defined hierarchy of relations:

The simplest kind, which I call *core consciousness*, provides the organism with a sense of self about one moment—now—and about one place—here. The scope of core consciousness is the here and now. Core consciousness does not illuminate the future, and the only past it vaguely lets us glimpse is that which occurred in the instant just before. There is no elsewhere, there is no before, there is no after. On the other hand, the complex kind of consciousness, which I call *extended consciousness* and of which there are many levels and grades, provides the organism with an elaborate sense of self—an identity and a person, you or me, no less—and places that person at a point in individual historical time, richly aware of the lived past and of the anticipated future, and keenly cognizant of the world beside it. . . .

When we think of the glory that is consciousness, and when we consider consciousness as distinctively human, we are thinking of extended consciousness at its zenith. And yet, as we shall see, extended consciousness is not an independent variety of consciousness: on the contrary, it is built on the foundation of core consciousness. The fine scalpel of neurological disease reveals that impairments of extended consciousness allow core consciousness to remain unscathed. By contrast, impairments that begin at the level of core consciousness demolish the entire edifice of consciousness: extended consciousness collapses as well. (Ibid., pp. 16–17)

The scalpel of neurology also suggests that core consciousness is a necessary condition (although not sufficient) for the emergence of extended consciousness: the epileptic patient, for example, lost his consciousness *in toto* because of the impairment of core consciousness. Just as consciousness is not a monolith, so too with the self:

Incidentally, the two kinds of consciousness correspond to two kinds of self. The sense of self which emerges in core consciousness is the *core self*, a transient entity, ceaselessly re-created for each and every object with which the brain interacts. Our traditional notion of self is linked to the idea of identity and corresponds to a nontransient collection of unique facts and ways of being which characterize a person. My term for that entity is *autobiographical self*. The autobiographical self depends on systematized memories of situations in which core consciousness was involved in the knowing of the most invariant characteristics of an organism's life—who you were born to, where, when, your likes and dislikes, the way you usually react to a problem or conflict, your name, and so on. (Ibid., p. 17)

The absence crisis shows the basic role of core consciousness and of the emergence of a core self, while other cases of impaired extended consciousness and autobiographical self show that the primitive core can still support a sense of the self even in dramatic situations. This means, for Damasio, that core conscious-

ness and the core self have to be the main targets of scientific research, because a theoretical account focused on higher-level cognitive processes will depend on an account of these more primitive phenomena.

The case of David makes this point quite well. This patient, described by Damasio, has an extensive damage in the areas called hippocampus, amygdala, and prefrontal lobes, resulting in a total loss of the power to produce new memories and, consequently, to learn new abilities and facts. Notwithstanding this serious impairment of extended consciousness, David's core self remains untouched and still capable of producing a primitive, emotive "knowledge" driving his actions.

In a controlled experimental setting, David interacted for two weeks with three different people: the "good guy," kind and educated, the "neutral guy," and the "bad guy," boring and uneducated. After some time, the doctors gave David some pictures of the guys to examine. He did not recognize any of them. But asked to answer which of them he thought was his friend, he selected the good guy's picture with a frequency of 80 percent, despite being unable to state the reasons of his choice.

The cases of David and of the absence crisis highlight a fundamental feature of consciousness: it produces in the organism the feeling of what happens and the sense that it is happening to that organism, from its individual point of view. The point is well illustrated in the following anecdote:

David was being brought to a bad-guy encounter and as he turned into the hallway and saw the bad guy awaiting him . . . he flinched, stopped for an instant, and only then allowed himself to be led gently to the examining room. I picked up on this and immediately asked him if anything was the matter, if there was anything I could do for him. But, true to form, he told me that, no, everything was all right—after all, nothing came to his mind, except, perhaps, an isolated sense of emotion without a cause behind that emotion. I have no doubt that the sight of the bad guy induced a brief emotional response and a brief here-and-now feeling. However, in the absence of an appropriately related set of images that would explain to him the cause of the reaction, the effect remained isolated, disconnected, and thus unmotivated. (*Ibid.*, p. 46)

From these cases emerges a well-defined picture of the embodied nature of consciousness, translated by Damasio into a theoretical proposal: an organism becomes conscious when the brain develops a set of devices representing the organism, the body, in interaction with the world. This self-regulative activity is what Damasio calls the proto-self:

I have come to conclude that the organism, as represented inside its own brain, is a likely biological forerunner for what eventually becomes the elusive sense of self. The deep roots for the self, including the elaborate self which encompasses identity and personhood, are to be found in the ensem-

ble of brain devices which continuously and *nonconsciously* maintain the body state within the narrow range and relative stability required for survival. These devices continually represent, *nonconsciously*, the state of the living body, along its many dimensions. I call the state of activity within the ensemble of such devices the *proto-self*, the nonconscious forerunner for the levels of self which appear in our minds as the conscious protagonists of consciousness: core self and autobiographical self.” (Ibid., p. 22)

In the context of this close connection between the emergence of consciousness and the neural processes involved in the self-regulation of the biological processes, Damasio puts forward a hypothesis about the evolutionary advantage of consciousness. The proto-self can place on the organism a sense of “ownership” with respect to its actions, thoughts, and mental states. This would be the ground of the wellspring of creativity:

If you find the connection between life and consciousness surprising, consider the following. Survival depends on finding and incorporating sources of energy and on preventing all sorts of situations which threaten the integrity of living tissues. . . . without actions, organisms such as ours would not survive since the sources of energy required for renewing the organism’s structure and maintain life would not be found and harnessed to the service of the organism, never mind staving off environmental dangers. But on their own, without the guidance of images, actions would not take us far. Good actions need the company of good images. Images allow us to choose among repertoires of previously available patterns of action and optimize the delivery of the chosen action. . . . We can pick and choose the most appropriate and reject the bad ones. Images also allow us to invent new actions to be applied to novel situations and to construct plans for future actions—the ability to transform and combine images of actions and scenarios is the wellspring of creativity.

If actions are at the root of survival and if their power is tied to the availability of guiding images, it follows that a device capable of maximizing the effective manipulation of images in the service of the interests of a particular organism would have given enormous advantages to the organisms that possessed the device and would probably have prevailed in evolution. Consciousness is precisely such a device. (Ibid., pp. 23–24)

We can reasonably say that the interrelations highlighted by Damasio among life, consciousness, sense of self, and evolutionary advantage, with the rise of creativity, responsibility, and ability of deliberate planning, are not only a confirmation of Searle’s philosophy of mind, but also indication that Searle’s theory is a good framework to account for these achievements and hypotheses.

We can remember that Damasio's thesis on the evolutionary and causal power of consciousness corresponds to that worked out by Searle in *The Rediscovery of the Mind* (1992), where he sees in the rising of creativity the kind of evolutionary advantage deriving from the emergence of consciousness in the natural world.

The complex logical relations structuring mind described in Searle's philosophy of mind justify this naturalized view of the evolutionary role of consciousness and, among these logical structures, the causal self-referentiality of perception and action plays a crucial role. The logical form of causal self-reference gives us a logical tool to understand how a prelinguistic organism endowed with basic biological powers, like perception and action, can develop a preconceptual ability of self-recognition as the deep root of a more complex sense of the self.

Speaking more generally, we can say that these recent developments in the field of neuroscience find in biological naturalism a framework that can account for them because it makes a non-reductive and realist view of mind compatible with the biological embodiment of consciousness in the natural world. In this sense, a desubstantialized view of mind also makes it possible, in virtue of the refutation of conceptual dualism, that naturalism and non-reductionism are compatible and that the evolutionary and causal role of consciousness in the life of the organism is intelligible. With consciousness, an organism acquires the decisive tool, which allows him to discover that he is an "I."

5. Subjective Ontology and the Construction of a Sense of the Self: The Interdependence of Holism, Ontology of Consciousness, and Mental Causation

With his book *Rationality in Action* (2001), Searle made a contribution to the debate on the self, working out a formal notion of the self as an answer to two different problems. Concerning the theory of rationality, how could we explain the causal power of reasons under the presupposition of the freedom of the will? With regard to the more general problem of the ontology of the mind, how could we define the concept of "I," or "subject," as it occurs in the definition of the subjective ontology of mental states?

Let us start the analysis with the first problem. Searle describes the logic underlying rational explanations of actions as involving a *sui generis* causality, different from physical causation because the antecedents of actions are not causally sufficient for the determination of the action itself. A gap stands between the action and its antecedents (beliefs, desires, intentions): the antecedent is not sufficient to causally determine the consequent.

But rational explanations of actions are, according to Searle, forms of causal explanations, since:

the normative constraints on the explanation of why an action occurred . . . do not remove the causal constraint that an explanation of why the agent did it must state the reasons that were effective with the agent. (Ibid., p. 112)

The problem then becomes, how we can account for the causal powers of reasons in the production of intentional actions if, in such cases, we do not deal with determinant causation.

In this context, Searle introduces the formal notion of the Self as an entity defined by features such as consciousness, power of acting on reasons under the constraints of rationality, and taking responsibility for actions. We are placing some constraints on an entity *X* that it must satisfy to be a Self, without “materially” saying what this “entity” is:

[W]e have to suppose that in addition to the “bundle of perceptions,” as described by Hume, there are certain formal constraints on the entity that makes the decisions and carries out the actions. We have to postulate a rational self or agent that is capable of acting freely and capable of assuming responsibility for actions. . . . To be more precise, in order to account for free, rational actions, we have to suppose there is a single entity *X* such that *X* is conscious . . . , *X* persists through time, *X* formulates and reflects on reasons for action under the constraints of rationality, *X* is capable of deciding, initiating and carrying out actions under the presupposition of freedom, and (already implicit in what I have said), *X* is responsible for at least some of its actions. (Searle, 2004, pp. 294–295)

Even with this “weak” formulation, we can face many objections raised against other realist views of the self as an explanation of rational actions. So for example, Searle’s view is not subject, unlike those worked out by Christine M. Korsgaard and Roderick M. Chisholm, to the critique of conceiving the self as a sort of special link in the causal chain going from the antecedents to the action. Searle’s Self does not fill the gap: it instead acts under the constraints of the gap. The Self, on this view, is the animal agent acting to make its reasons effective in actions under the constraints of rationality and free will and driving the elements in the unified field of consciousness from one stage to the next. The problem here is not that of waiting for the reasons that produce the action by themselves: we have to postulate the Self as the agent engaged in these operations.

The notion of the Self as the “immanent” or “transeunt” cause (Chisholm’s terminology) does not work for two reasons (Searle, 2001, pp. 82–83). First, in causal explanations, the causally relevant terms are not objects and entities but their properties and features, or an event implying properties and features. We cannot say that Bill caused the fire; instead the cause was that Bill distractedly put the lighter too close to the leaves.

Analogously, in the case of the explanation of an action the self does not explain anything unless we say that the agent acted that way because he had such

and such reasons. But this does not close the gap; it opens it again. Searle holds that we have no reason to close the gap, since the problem of filling the gap is itself ill posed. He says:

The fact that we are inclined to suppose that all explanations must fit a pre-conceived model of billiard ball causation is a limitation on our Background sensibility that I am now trying to overcome. (Ibid., p. 88)

If Searle's diagnosis is right, then an explanation of rational action does not have to state causally sufficient conditions for the action:

The causal gap does not imply an explanatory gap. The question, "Why did you do that?" does not ask: what causes were sufficient to determine your action? but rather it asks: what reason(s) did you, as a rational self, act on? And the answer to that question explains not by showing how the act as a natural event was inevitable given the antecedent causes, but by *showing how a rational self operated in the gap*. In a Wittgensteinian tone of voice one wants to say: this is how the language game of explaining actions is played, and don't suppose it must be played according to the rules of the language game of explanations in classic mechanics. (Ibid., pp. 85–86)

The Self, then, is not immanent causation, but we have to postulate it in order to make rational action intelligible. This view appears to be immune to the traditional Humean charges concerning the hypostatization of a grammatical fiction and postulation of a ghost in the machine. We are not saying that the Self is a "thing," instead it is a formal requirement of the logical structure underlying the rational explanation of actions:

How can we be sure that the apparent requirement of the postulation of a self is not just a grammatical illusion foisted on us by the subject-predicate structure of the sentences? Are we not reifying something in order to have an object for the "I" to refer to . . . ? No. For the grammatical requirement is the same even in cases where I am not doing anything . . . you do not capture the active feature of the decision by saying this sequence of experiences now includes a decision, for the decision was something I made, an action on my part. . . .

But are we not postulating a homunculus who lives in the gap and makes our decisions for us? And does this not lead to an infinite regress? No, because we live in the gaps and make the decisions. (Ibid., p. 94)

Is this answer also sufficient for our second problem, the ontological one? Let us recall its terms: what is the subject as defined in the concept of subjective ontology? Searle's Self is defined by conditions like free choice between differ-

ent plans of action under the constraint of rationality, so that the solution can be useful to characterize the so-called “higher-order consciousness” (Edelman, 1989) or “extended consciousness” (Damasio, 1999), which is typically human because it involves a socially and linguistically structured self-consciousness. But the formal Self is not useful in characterizing the first-person ontology of mental states because subjective ontology is shared by us, in Searle’s view, with prelinguistic and non-rational creatures (Searle, 2002ff; see Dennett, 1996 for a different perspective).

One of the problems is that the concept of subjective ontology, although apparently well defined, is more difficult when we articulate its consequences. Pains, thoughts, emotions, and sensory experiences exist only insofar as subjects experience them. Moreover, this feature distinguishes the form of existence of mental phenomena from other natural phenomena, which exist independent of their being experienced: subjective ontology is the ground of Searle’s thesis of irreducibility, so that a deeper understanding of the concept is crucial in his theoretical perspective.

The question is this: Can we define, in a non circular way, the “I” mentioned in the definition of subjective ontology (Honderich, 2001)? Consider some possibilities: the “I” cannot be the unified field of consciousness because consciousness already has a subjective ontology. A more promising possibility is that the subject is a human or animal organism. But if an organism has agency (crucial for the existence of an organism in a biological world governed by selectionist criteria) the circularity problem returns. If acting is different from the mere happening of bodily movements, the relevant difference is that the first, but not the second, has ontologically subjective mental states as its internal causal basis.

But if we can understand the genesis of this primitive subjectivity, then we could account for the existence of a non-substantial subject of mental states that can develop a Self (in Searle’s sense of the word).

I think that we can give a possible answer to the question following the “conceptual” genesis of the Self within the framework of Searle’s philosophy of mind. Mental phenomena are, according to Searle, ontologically irreducible processes caused by and realized in highly complex neurophysiologic systems (such as human and animal brains) according to the mechanism of bottom-up, no time gap causation. Mind as such, as subjective and qualitative, is a level of the organization of the brain. Mindful living beings are a part of nature qua mindful, where we have to interpret this qualification against conceptual dualism, according to which the mental as such is suspect with respect to the (materialistically interpreted) scientific worldview.

Non-reductive embodiment of mind within the biological life of an organism grounds the crucial contribution that consciousness gives, under the causal and evolutionary aspect, to the self-regulation of organism-environment interactions. This contribution is the power of consciousness to give the organism behavioral creativity and flexibility.

So far, biological naturalism states two constraints on the concept of mind. First, its embodiment: mind is caused and realized by organic systems, or by physical systems complex enough to produce consciousness. Second, mind causally acts in the regulation of the behavior of the organism.

In addition to the constraints of embodiment and causal power of mind, we have two other theses: the interdependence between qualitateness, subjectivity, and holistic unity and the thesis of the holistic structure of intentionality.

The first thesis says that conscious mental states are qualitative because they are ontologically subjective (since their existence depends on the subject experiencing them) and that in turn, this qualitative subjectivity will articulate its states in a holistic unity. The different states will be such only within a unified field the internal relations of which will be partly constitutive of the different states. The second thesis states that every intentional state determines its conditions of satisfaction once it assumes a position within a Network of other intentional states and against a Background of mental, preintentional skills that constitute the pragmatic structure of the Network. Since the structure of consciousness is a manifestation of the logical structure of intentionality (Searle, 2004, p. 174) we could say that the first thesis is justified by the second.

At this point, we have an organism with qualitative, subjective, holistically structured states that, by virtue of the bootstrapping effect produced by the interaction of prelinguistic intentionality with Background abilities and linguistic structures (learned and developed in an institutional context), evolves higher-level cognitive abilities defining it as a Self.

But what is it, to begin with, that makes the organism capable of being a self in the sense of subject occurring in the definition of subjective ontology? What is the genesis of the rational Self, engaged in carrying out its reasons? I believe that a theoretical-interpretive hypothesis about the genesis of the Self could start with an account of the relations between embodiment, causal powers of consciousness, and holism. We should analyze the link relating holism and mental causation in the genesis of subjectivity under the constraint of embodiment.

Searle's analysis of the logical structure of intentionality accounts for the power of consciousness to creatively structure the behavior of the organism. The structure underlying this power is that the organism develops its Network of intentional states from biologically primitive mental abilities with a causally self-referential component (perception and action). This element introduces a primitive form of indexicality "in the head" of the organism before it develops language and the ability of manipulating the concept of "I."

Perception and action have a causally self-referential element as an essential feature of the logical structure of their contents. The organism capable of action and perception experiences itself as causing or caused by the rest of the conditions of satisfaction specified in the content. This primitive form of self-reference, as Searle writes in his reply to Tyler Burge's objections (1991), does not require that the animal can make explicit the contents of actions and perceptions in the form of second-order thoughts or speech acts. This primitive self-reference does not require

self-consciousness and does not require inferentially articulated conceptual abilities (Searle, 1991c, pp. 183–184; 1991e; 1991f, pp. 228–229). Action and perception have this structure and the availability in conceptual terms is not necessary to guarantee their functioning.

Using Damasio's words, we could say not only that the conscious field is unified because it is embodied, but also that the embodiment requires that the holistic structure of mind acts *for the benefit* of the organism. The construction of the field from causally self-referential elements appears to be a good starting point for the construction of the organism as a Self. The way the conscious mind can structure the behavior of the organism consists in enabling it to *know*, or better yet, to primitively *feel* that the action is *its* action, that the consequences will give or not give *it* a benefit. The organism will have to "know" (again, in a quite primitive sense) that the interaction with the environment finds in that very organism one of the terms of the relation. When the organism sees the object, or the scene, it sees the environment from its point of view and it has to know that it is *its* point of view: it is that very organism, for example, which has to run, because it sees the prey or the predator.

For example, the ape and the prelinguistic child can successfully do the "mark test." The mark test is the typical experiment used to test the ability of self-recognition in animals and human beings. The subject has a mark on its face and if the subject, posed in front of the mirror, uses it to clean its face and does not try to clean the mirror, this means that he "thinks" that the picture is "its" picture, and not that of another creature. The subjects can coordinate action and perception in front of the mirror with respect to their mirror image. This apparently means that the subjects' ability to recognize themselves in the mirror image does not depend on or imply the development of conceptual or linguistic structures. The sense of the self is primitive with respect to both. What I am saying is that the sense of the self is produced by the primitive forms of causal self-reference and indexicality that we get in perception and action.

This solution has some advantages over other accounts of the self, such as those above examined. The idea of the genesis of a rational Self in the biologically and logically primitive forms of causal self-reference shows the intimate continuity between human beings and the rest of the natural world, against the supposed discontinuity implied by Wilfrid Sellars, Donald Davidson, and John McDowell. Moreover, we did not find a definition of mind in non-mental terms, but our goal was not to overcome the circularity problem in defining subjectivity with a reductionist approach. We have just to see perception and action as biological and mental primitive processes, with the implication that an organism endowed with these processes has a minimal sense of the self developed through causal self-referentiality. If we see perceptual experiences and intentions as biological processes caused by and realized in human and animal brains we will not interpret the "I" as a thing (so we will not give voice to neo-Humean objections) and we will not consider it as a sort of useful fiction (Dennett). We will not eliminate the subject, unlike the extended mind approach, because the embodi-

ment criterion and its consequences are crucial for the development of the self as a fundamental building block of the account.

Also, these considerations on the construction of a sense of the self from the primitive forms of causal self-reference allow for a reinterpretation of Searle's theory and its implications of a close interdependence among the ontology of mind, mental causation, and holistic structure. If Searle argues for the interdependence of the essential features of consciousness (qualitativeness, subjectivity and holistic unity), we can add that this interdependence finds its origin in the more general holistic structure of the representational Network of mental states. This means not only that the Network contains causally self-referential elements, but also that intentional states lacking this feature are elements whose causal self-reference is in some way "bleached out," as if the occurrence in causal relations was somehow "inherited" by belief and desire from the primitive forms of intentionality (Searle, 1983, p. 96).

On the other hand, if the interpretation I put forward is correct, my thesis can be better stated in this form: we should conceive of consciousness as a process creating holistically unified fields the biologically primitive elements of which produce a space of self-referentiality (the subject). The functional role of this conscious subjectivity is the creative management of mind-world causal transactions. The underlying structure of this role is made possible by the generation of that primitive sense of the self, which makes intentional states subjective.

This interpretation of the consequences of holism and mental causation is along the same lines of Damasio's neurobiological theory of consciousness. The genesis of the self comes from the forms of causal self-referentiality within the holistic Network and under the constraint of embodiment. If we follow this line of reasoning, we can conclude that the embodiment and causal power of consciousness have a heuristic value for the construction of a neurobiological theory of consciousness and of the self.

CONCLUSION

Starting from the 1980s, the cultural landscape saw a growing number of scientific and philosophical studies of consciousness with many books and monographic issues of journals devoted to this issue (see for example the essays collected in the monographic number 79, 2001 of *Cognition*, which includes Block, 2001; Dennett, 2001, Parvizi and Damasio, 2001; Dehane and Naccache, 2001).

For the most part, these studies inherited a wider philosophical and scientific background that espouses the desubstantialization of mind, the thesis that mental states and abilities are not things or substances, but processes non-reductively embodied as part of the biological makeup of organisms and responsible for the mediation of the organism-environment relationships. The desubstantialization of mind amounts to an attempt to understand mental processes as a part of the biological order of the natural world without leaving out their essential features, that John R. Searle describes as qualitateness, subjectivity, and unity, and to show how these features are closely related to their embodied nature. Consciousness is qualitative because it is subjective and this qualitative subjectivity finds in its embodiment the reasons for its unity, and its being situated in an individual perspective.

Notwithstanding the shift towards this richer and more complex view of mind, which I analyzed in the present work with reference to the works of Locke and Searle, most of the debate in philosophy of mind appears to remain deeply linked to a Cartesian conceptual framework. Materialists and dualists cannot see subjectivity as such as a part of nature as such.

Authors such as Richard Rorty, Daniel C. Dennett, and Jaegwon Kim often start with a verificationist and objectivist picture of science. Next they construct models of the ontology and functioning of mind in which consciousness does not play any role in the production of behavior. Their claim that this way of demystifying the mind is a scientific and philosophical achievement, against a Cartesian and spiritualistic philosophical mentality, is often offered as the reason grounding this materialistic move.

On the other side, authors like Sir Karl Raimund Popper, Thomas Nagel, Colin McGinn, and David J. Chalmers grant the existence of conscious phenomena and their importance for the ordinary self-conception of us as real agents. But some from that group also describe the physical and the mental as separated worlds (Popper), or state that phenomenal properties are ontologically independent of the physical and causally inert (Chalmers).

The result is that the debate appears to force us to choose between materialism, that appears incapable of “saving the phenomena,” and a range of skeptical or neodualistic solutions that reject the idea that conscious beings are a part of the natural world.

On the contrary, Searle’s philosophy of mind shows that antireductionism and naturalism are coherently conjugable within a unified theory of the ontology

and functioning of mind. Overcoming the traditional monism-dualism dilemma is necessary if we want to work out reliable solutions to the main problems in current philosophical debate, such as the problems of consciousness, mental causation, and the self, because these problems find their roots in that dilemma. Searle's explicit critique aims at getting rid of conceptual dualism, the set of theoretical assumptions (mainly the principle of mutual ontological exclusion between the mental and the physical) that systematically would make philosophers blind to the possibility that mind, with its subjectivity, is part of nature.

Searle's biological naturalism conceives of mind as a process or macrofeature realized in the physical structure of the brain at a higher and ontologically irreducible level of its organization. In the context of Searle's thought, mind is therefore "part of our biological natural history" as other biological processes (1992, p. 1). Because mind is embodied as part of the biological makeup of some organisms, it can play a role in the modulation of the behavior of the organism. A being with a conscious mind, Searle says, can have a more rich, flexible, and discriminative behavior than that of an unconscious mechanism (*ibid.*, pp. 106–109).

In this work, I reconstructed the philosophical debate on the mind-body problem and compared Searle's perspective with Dennett and Chalmers's theories of consciousness and with Kim's thinking on the problem of mental causation.

Dennett's core thesis is that phenomenal consciousness does not exist because no Cartesian Theater, no place in the brain exists at which the information converges to become conscious. Instead, the brain has chaotic processes in the brain, all of them objective, third-person, and responsible for information processing on multiple channels.

But Dennett's argument does not work because it is question-begging in that it presupposes a reductive one-to-one correspondence between conscious experience and brain states. Also, Dennett never explains why the reality of consciousness would depend on the existence of the Cartesian Theatre. To take a non-tendentious example, nobody would deny the real existence of language, even though no "language headquarters" exists in the brain.

Equally tendentious appears the equivalence often proposed by Dennett between the Cartesian concept of idea and the contemporary concept of qualia. But according to Searle, Tim Crane, and others, the qualia jargon is at best ambiguous and no necessary connection exists between the concept of consciousness and the Cartesian dogmas correctly criticized by Dennett.

So Dennett's position does not state valid reasons for the elimination of phenomenal consciousness.

David Chalmers's main thesis is that we have to consider consciousness and phenomenal properties as ontologically independent of the physical because consciousness does not logically supervene on physical facts. Maybe, Chalmers says, we could hope to find the contingent bridge laws that associate consciousness to the physical world. But still, even though we could understand *how*, we will never understand *why* consciousness exists in the real world as opposed to a hy-

pothetical zombie world. Moreover, Chalmers argues, this could lead us to the acceptance of some variety of epiphenomenalism.

Chalmers's dualistic and epiphenomenal conclusions are based on weak arguments. The zombie argument used to argue for property dualism appears to be question-begging in that it presupposes dualism instead of demonstrating it. If we can clearly imagine a world physically identical to ours, but with a zombie twin for every conscious being in the actual world, this would prove that no logical connection exists between consciousness and the physical.

The possibility of dualism is logically coherent only because Chalmers's argument assumes from the beginning the logical coherence of a zombie world. For example, even though a vitalist could conceive, during the nineteenth century, life as not logically supervenient on physics, nowadays Chalmers says that biology logically supervenes on physics: no logical room exists for biological facts to vary once the physical facts are determined. Why should things be different for consciousness? Why could a neurobiological theory not explain how and why consciousness arises in the real world?

Dennett's and Chalmers's positions are worked out, with more or less explicit awareness, within the theoretical framework of conceptual dualism. My goal in examining and assessing these theories has been to show that Searle's biological naturalism is immune to the problems of these theories because it puts forward an alternative theoretical framework that gives intelligibility to a view of consciousness as a natural, subjective, qualitative, and holistically unified phenomenon without the constraints of conceptual dualism.

The examination of the debate between Kim and Searle on the dilemma of causal exclusion showed that the supposed incompatibility of antireductionism, causal closure of the physical, and causal efficacy of mental states as stated by Kim's argument presupposes an unjustified interpretation of antireductionism.

If we conceive of mind as a system macrofeature realized at a higher level of the organization of the system, and we accept the logic of Kim's dilemma for mind, then the causal exclusion argument has to be generalized against the causal efficacy of any property, process, or state at a level higher than that of the basic physics. But this position would be too implausible to be taken seriously. It appears instead to prove that the hidden premise articulating the dilemma of causal exclusion is still conceptual dualism.

Searle can answer Kim's argument with a detailed analysis of the reasons (the existence of bottom-up, no time gap causation) why, when we talk about causal efficacy of the mental, we are not talking about something happening in opposition, in addition, or in parallel to the causal efficacy of the physical system which embodies and explains mind. Instead, we are describing the system at different levels of organization, each one of them explained by the most fundamental level, but not reducible to it.

Searle also works out two further theses. Consciousness is a device responsible for the creative management of the organism-environment relations. In addition, qualitateness, subjectivity, and holistic structure of mind are interdependent.

The analysis of Searle's theory of intentionality pointed out that these theses are grounded in the more general logical structure of intentional states and in the complex relationships among holistic Network, preintentional Background abilities, and causal self-referentiality in perception and action.

The critical comparison between this theory and the current holistic paradigmatic turn in neuroscience, which I followed with reference to Gerald M. Edelman and Giulio Tononi's theory of consciousness, showed that Searle's philosophy of mind has specific conformations in current scientific research. Edelman and Tononi, like Searle, conceive of consciousness as a subjective and holistic process irreducibly embodied in an organism, where the holistic nature of the process implies that we cannot divide it in independent elements and reduce it to a single brain place or process. For them, the target of scientific research is the whole integrated and differentiated subjective conscious field as a level of the organization of the brain, the level of a "functional cluster" composed of many other holistically related processes.

For Edelman and Tononi, like for Searle, the subjective nature of consciousness cannot be reduced to objective items, but this does not mean that we cannot scientifically study consciousness or that we have to postulate a Cartesian Theater to explain subjectivity.

Damasio's theory gives us some further specification to the view of consciousness as a subjective and holistic process. The holistic field is unified because it is embodied. Embodiment in turn requires that consciousness act for the benefit of the organism.

Consciousness performs a causal and evolutionary role in the life of the organism. Damasio, like Searle, defines this role as the introduction of flexibility and creativity in the behavior of the organism. Consciousness can perform this role because it is rooted in the proto-self of the organism as represented in the self-monitoring brain devices, the ancestors of the sense of the self typically present in the field of consciousness.

The evolutionary function of consciousness, then, consists in giving the organism a self, a sense of being involved in interactions with the environment, a sense of the fact that the consequences of that interaction will have effect on that very organism.

We can find in Searle's book, *Intentionality*, the origin of the idea of a proto-self with an evolutionary role in the emergence of the higher-level cognitive abilities and of the creative management of mind-world relations. The representational Network contains biologically and logically primitive intentional states, the contents of which are causally self-referential. The contents of perception and action refer to the state itself as caused by or causing its conditions of satisfaction. When an organism has the primitive mental and biological abilities of perceiving and acting, it can experience itself as a self, as a subject that can change or be changed by the world.

If this analysis is correct, we can reasonably conclude that Searle's naturalistic and non-reductive philosophy of mind integrates the desubstantialization of

mind within a framework capable of overcoming the conceptual limitations of monism and dualism. The result is a model of mind that can fruitfully interact with science and efficaciously face the traditional challenges coming from the problems of consciousness and mental causation.

WORKS CITED

- Alexander, Samuel. (1920) *Space, Time, and Deity*. 2 vol., London: Macmillan.
- Apel, Karl O. (1991) "Is Intentionality more Basic than Linguistic Meaning?" In Lepore and Van Gulick, *John Searle and His Critics*, pp. 31–55.
- Armstrong, David M. (1991) "Intentionality, Perception, and Causality: Reflections on John Searle's *Intentionality*." In Lepore and Van Gulick, *John Searle and His Critics*, pp. 149–158.
- Atherton, Margaret, ed. (1999) *The Empiricists: Critical Essays on Locke, Berkeley, and Hume*. Lanham: Rowman and Littlefield.
- Bayne, Tim, and David J. Chalmers. (2003) "What is the Unity of Consciousness?" In Cleeremans, *The Unity of Consciousness*, pp. 23–58.
- Bechtel, William. (1988) *Philosophy of Mind: An Overview for Cognitive Science*. Hillsdale: Erlbaum.
- Bennett, Jonathan. (1999) "Locke's Philosophy of Mind." In Chappell, *The Cambridge Companion to Locke*, pp. 89–113.
- Block, Ned. (1978) "Troubles with Functionalism." In C. Wade Savage, ed. *Minnesota Studies in the Philosophy of Science*, vol. 9, *Perception and Cognition*, Minneapolis: University of Minnesota Press, pp. 261–325.
- . (1993) "Review of Dennett, *Consciousness Explained*," *Journal of Philosophy*, 90:4, pp. 181–193.
- . (1997) "On a Confusion about a Function of Consciousness," Block, Flanagan, and Güzelde, *The Nature of Consciousness*, pp. 375–415; originally published in *Brain and Behavioral Sciences*, 18:2 (1995), pp. 227–287.
- . (2001) "Paradox and Cross-Purposes in Recent Work on Consciousness," *Cognition*, 79, pp. 197–219.
- , O. Flanagan, and Güven Güzelde, eds. (1997) *The Nature of Consciousness*. Cambridge, Mass.: MIT Press.
- Bocchi, Gianluca, and Mauro Ceruti, eds. (1992) *La sfida della complessità (The Challenge of Complexity)*. Milano: Feltrinelli.
- Bodei, Remo. (2003) *Destini personali. L'età della colonizzazione delle coscienze (Personal Destinies: The Age of the Colonization of Consciousnesses)*. Milano: Feltrinelli.
- Bornstein, Robert F., and Thane S. Pittman, eds. (1992) *Perception without Awareness*. New York: Guilford Press.
- Broad, Charles D. (1925) *The Mind and its Place in Nature*. London: Routledge & Kegan Paul.
- Brook, Andrew, and Don Ross, eds. (2002) *Daniel Dennett*. Cambridge, UK: Cambridge University Press.
- Burge, Tyler. (1991) "Vision and Intentional Content." In Lepore and Van Gulick, *John Searle and His Critics*, pp. 195–213.
- . (1993) "Mind-Body Problem and Explanatory Practice." In Heil and Mele, *Mental Causation*, pp. 97–120.
- Carli, Eddy. (1997) *Cervelli che parlano (Speaking Brains)*. Milano: Mondadori.
- Chalmers, David J. (1996) *The Conscious Mind: In Search of a Fundamental Theory*. Oxford: Oxford University Press.

- . (1997) "On Consciousness and the Philosophers." <http://www.consc.net/book/searle-response2.html>.
- , ed. (2002) *Philosophy of Mind: Classical and Contemporary Readings*. Oxford: Oxford University Press.
- , and John R. Searle. (1997) "An Exchange with David Chalmers." In Searle, *The Mystery of Consciousness*, pp. 163–167; originally published as "Consciousness and the Philosophers: An Exchange," *New York Review of Books*, 44:8, 15 May 1997.
- Changeux, Jean-Pierre. (2004) *The Physiology of Truth: Neuroscience and Human Knowledge*. Translated by M. B. DeBevoise. Cambridge, Mass.: Harvard University Press.
- Chappell, Vere, ed. (1999) *The Cambridge Companion to Locke*. Cambridge, UK: Cambridge University Press.
- Chomsky, Noam. (2000) *New Horizons in the Study of Language and Mind*. Cambridge, UK: Cambridge University Press.
- Churchland, Paul M. (2002) "The Rediscovery of Light." In Chalmers, *Philosophy of Mind*, pp. 362–371; originally published in *Journal of Philosophy*, 93 (1996), pp. 211–228.
- Clark, Andy. (1997) *Being There*. Cambridge, Mass.: MIT Press.
- , and David J. Chalmers. (2002) "The Extended Mind." In Chalmers, *Philosophy of Mind*, pp. 643–651; originally published in *Analysis*, 58:1 (1998), pp. 7–19.
- Cleeremans, Axel. (2003) *The Unity of Consciousness: Binding, Integration, and Dissociation*. Oxford: Oxford University Press.
- Crane, Tim. (1995) "The Mental Causation Debate," *Proceedings of the Aristotelian Society*. Supplementary Volume, 69, pp. 211–236.
- . (2001a) "The Significance of the Emergence." In Carl Gillett and Barry Loewer, *Physicalism and its Discontents*. Cambridge, UK: Cambridge University Press.
- . (2001b) *Elements of Mind: An Introduction to the Philosophy of Mind*. Oxford: Oxford University Press.
- . (2003) *The Mechanical Mind: A Philosophical Introduction to Minds, Machines and Mental Representations*. 2nd ed. London: Routledge.
- Crick, Francis. (1994) *The Astonishing Hypothesis: The Scientific Search for the Soul*. New York: Scribner.
- , and Christof Koch. (1997) "Towards a Neurobiological Theory of Consciousness." In Block, Flanagan, and Güzelidere, *The Nature of Consciousness*, pp. 277–292; originally published in *Seminars in the Neurosciences*, 2 (1995), pp. 263–275.
- Dalbhon, Bo, ed. (1993) *Dennett and His Critics*. Oxford: Blackwell.
- Damasio, Antonio. (1994) *Descartes' Error: Emotion, Reason and the Human Brain*. New York: Gosset/Putnam Press.
- . (1999) *The Feeling of What Happens*. San Diego: Harcourt.
- , and Hanna Damasio. (1996) "Making Images and Creating Subjectivity." In Linas and Churchland, *The Mind-Brain Continuum*, pp. 19–28.
- Davidson, Donald. (1980a) *Essays on Actions and Events*. New York: Oxford University Press.

- . (1980b) "Mental Events." In Davidson, *Essays on Actions and Events*, pp. 207–228; originally published in Lawrence Foster and Joe William Swanson, eds. *Experience and Theory*. Amherst, Mass.: University of Massachusetts Press, 1970, pp. 79–101.
- . (1984a) *Inquires into Truth and Interpretation*. Oxford: Oxford University Press.
- . (1984b) "Thought and Talk." In Davidson, *Inquires into Truth and Interpretation*, pp. 155–170.
- Dehane, Stanislas, and Lionel Naccache. (2001) "Toward a Cognitive Neuroscience of Consciousness: Basic Evidence and a Workspace Framework," *Cognition*, 79, pp. 1–37.
- Dennett, Daniel C. (1969) *Content and Consciousness*. London: Routledge & Kegan Paul.
- . (1980) "The Milk of Human Intentionality," *Behavioral and Brain Sciences*, 3:3, pp. 428–430.
- . (1982) "Comments on Rorty," *Synthese*, 53, pp. 349–356.
- . (1987) *The Intentional Stance*. Cambridge, Mass.: MIT Press.
- . (1991) *Consciousness Explained*. Boston: Little Brown and Company.
- . (1993a) "Review of Searle, *The Rediscovery of Mind*," *Journal of Philosophy*, 60:4, pp. 193–205.
- . (1993b) "Précis of *Consciousness Explained*," *Philosophy and Phenomenological Research*, 53:4, pp. 889–892.
- . (1993c) "The Message Is: There Is No Medium," *Philosophy and Phenomenological Research*, 53:4, pp. 889–931.
- . (1995) "Dennett, Daniel C." In Guttenplan, *A Companion to the Philosophy of Mind*, pp. 236–244.
- . (1996) *Kinds of Minds*. New York: Harper & Collins.
- . (1997) "*La coscienza: un eterno enigma filosofico*" ("Consciousness: An Eternal Philosophical Enigma.") In Carli, *Cervelli che parlano (Speaking Brains)*, pp. 63–85.
- . (2001) "Are We Explaining Consciousness Yet?," *Cognition*, 79, pp. 221–237.
- . (2002) "Quining Qualia." In Chalmers, *Philosophy of Mind*, pp. 226–246; originally published in Anthony J. Marcel and Edoardo Bisiach. *Consciousness and Contemporary Science*. Oxford: Oxford University Press, 1988, pp. 42–77.
- . (2005) *Sweet Dreams: Philosophical Obstacles to a Science of Consciousness*. Cambridge, Mass.: MIT Press.
- , and John R. Searle (1997) "'The Mystery of Consciousness': An Exchange." In Searle, *The Mystery of Consciousness*, pp. 115–120; originally published in *The New York Review of Books*, 21 December, 1995, pp. 83–85.
- Descartes, Renè. (1989) *The Passions of the Soul*. Translated by Stephen H. Voss. Indianapolis, Ind.: Hackett.
- . (1993) *Meditations on First Philosophy: In Which the Existence of God and the Distinction of the Soul from the Body Are Demonstrated*. Translated by Donald A. Cress. Indianapolis, Ind.: Hackett.
- . (1998) *Discourse on the Method for Conducting One's Reason Well and for Seeking Truth in the Sciences*. Translated by Donald A. Cress. Indianapolis, Ind.: Hackett.
- Di Francesco, Michele. (1998) *L'io e i suoi sé: Identità personale e scienze della mente (The I and Its Selves: Personal Identity and Sciences of Mind)*. Milano: Cortina.

- . (2000) *La coscienza (Consciousness)*. Roma and Bari, Italy: Laterza.
- . (2002) *Introduzione alla filosofia della mente (Introduction to the Philosophy of Mind)*. 2nd ed. Milano: Carocci.
- . (2004) “‘Mi ritorni in ment’: *Mente distribuita e disunione del soggetto*” (“‘You Come Back Into My Mind’: *Distributed Mind and the Disunity of the Subject*”), *Networks*, 3–4, pp. 115–139.
- Di Lorenzo Ajello, Francesca. (2000) “*Giudizi di valore e pretese di validità: Implicazioni deontiche della speech act theory*” (*Value Judgments and Validity Claims: Deontic Implications of the Speech Acts Theory*”). In Salvatore Nicosia, ed. *Il Giudizio: Filosofia, teologia, diritto, estetica (Judgment: Philosophy, Theology, Right, Aesthetics)*. Roma: Carocci.
- . (2001) *Mente, azione e linguaggio nel pensiero di John R. Searle (Mind, Action, and Language in John R. Searle’s Thought)*. Milano: Franco Angeli.
- . (2003) “Over the Dichotomy between Cognitive Judgments and Value Judgments: Speech Acts and Commitments.” In Rosaria Egidi, Marcello Dell’Utri, and Mario De Caro, eds. *Normatività, fatti, valore (Normativity, Facts, Value)*. Macerata, Italy: Quodlibet.
- Di Lucia, Paolo, ed. (2003) *Ontologia sociale: Potere deontico e regole costitutive (Social Ontology: Deontic Power and Constitutive Rules)*. Macerata, Italy: Quodlibet.
- Edelman, Gerald M. (1989) *The Remembered Present: A Biological Theory of Consciousness*. New York: Basic Books.
- . (1992) *Bright Air, Brilliant Fire: On the Matter of the Mind*. London: Basic Books.
- . (2004) *Wider than the Sky: The Phenomenal Gift of Consciousness*. New Haven, Conn.: Yale University Press.
- , Giulio Tononi. (2000) *A Universe of Consciousness: How Matter Becomes Imagination*. New York: Basic Books.
- Elton, Matthew. (2003) *Daniel Dennett: Reconciling Science and Our Self-Conception*. Oxford: Polity Press.
- Fodor, Jerry A. (1987) *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. Cambridge, Mass.: MIT Press.
- . (1990) *A Theory of Content and Other Essays*. Cambridge, Mass.: MIT Press.
- . (2002) “Special Sciences (or: the Disunity of Science as a Working Hypothesis).” In Chalmers, *Philosophy of Mind*, pp. 126–135; originally published in *Synthese*, 28(1974), pp. 97–115.
- Fotion, Nick. (2000) *John Searle*. Teddington, UK: Acumen.
- Freeman, Walter. (2000) *How Brains Make Up Their Minds*. New York: Columbia University Press.
- , and Christine Skarda. (1991) “Mind/Brain Science: Neuroscience on the Philosophy of Mind.” In Lepore and Van Gulick, *John Searle and His Critics*, pp. 115–127.
- Gargani, Aldo. (1966) “Idea, mondo, e linguaggio in Th Hobbes e J. Locke” (“Idea, World, and Language in Th. Hobbes and J. Locke”), *Annali della Scuola Normale superiore di Pisa*, Series 2, 35, pp. 251–292.
- Gazzaniga, Michael. (1998) *The Mind’s Past*. Berkeley, Calif.: The University of California Press.

- Glaserfeld, Ernst, von. (1992) "Il complesso di semplicità" ("The Simplicity Complex"). In Bocchi and Ceruti, *La sfida della complessità (The Challenge of Complexity)*, pp. 103–111.
- Grewendorf, Gunther, and Georg Meggle, eds. (2002) *Speech Acts, Mind, and Social Reality. Discussions with John R. Searle*. Dordrecht, The Netherlands: Kluwer.
- Gulì, Sandro. (2007) *Elementi, sensazioni, e connessioni funzionali: La filosofia naturale di Ernst Mach (Elements, Sensations, and Functional Connections: Ernst Mach's Natural Philosophy)*. Milano: Unicopli.
- Guttenplan, Samuel. (1995) *A Companion to the Philosophy of Mind*. Oxford: Blackwell.
- Frege, Gottlob (1980) "On Sense and Reference." Translated by M. Black in *Translations from the Philosophical Writings of Gottlob Frege*, P. Geach and M. Black (eds. and trans.). 3rd ed. Oxford: Blackwell.
- Habermas, Jürgen. (1991) "Comments on John Searle: 'Meaning, Communication, and Representation.'" In Lepore and Van Gulick, *John Searle and His Critics*, pp. 17–29.
- Heil, John, and Alfred Mele, eds. (1993) *Mental Causation*. Oxford: Oxford University Press.
- Hodgson, David. (1994) "Why Searle has not rediscovered the Mind," *Journal of Consciousness Studies*, 1:2, pp. 264–274.
- Hofstadter, Douglas R. (1979) *Gödel, Escher, Bach: An Eternal Golden Braid*. New York: Basic Books.
- . (1980) "Reductionism and Religion," *Behavioral and Brain Sciences*, 3:3, pp. 433–434.
- . (1981) "'Reflections': Commentary on Searle, 'Minds, Brains, and Programs.'" In Hofstadter and Dennett, *The Mind's I*, pp. 373–382.
- . (2007) *I am a Strange Loop*. New York: Basic Books.
- , and Daniel C. Dennett. (1985) *The Mind's I*. New York: Basic Books.
- Holt, Lynn. (1999) "Metaphor, History, Consciousness: From Locke to Dennett," *The Philosophical Forum*, 30:3, pp. 187–200.
- Honderich, Ted. (2001) "Mind the Guff: A Response to John Searle," *Journal of Consciousness Studies*, 8:4, pp. 62–78.
- Hume, David. (1984) *A Treatise of Human Nature*. London: Penguin Books.
- Jackson, Frank. (1993) "Appendix A (For Philosophers)," *Philosophy and Phenomenological Research*, 53:4, pp. 899–903.
- . (1997) "What Mary Didn't Know." In Block, Flanagan, and Güzeldere, *The Nature of Consciousness*, pp. 567–570; originally published in *Journal of Philosophy*, 83 (1986), pp. 291–295.
- . (2002) "Epiphenomenal Qualia." In Chalmers, *Philosophy of Mind*, pp. 273–280; originally published in *Philosophical Quarterly*, 32 (1982), pp. 127–136.
- Jolley, Nicholas. (1999) *Locke: His Philosophical Thought*. Oxford: Oxford University Press.
- Kim, Jaegwon. (1993) *Supervenience and Mind*. Cambridge, UK: Cambridge University Press.
- . (1995) "Mental Causation in Searle's Biological Naturalism," *Philosophy and Phenomenological Research*, 55:1, pp. 189–194.
- . (1998) *Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation*. Cambridge, Mass.: MIT Press.
- . (2006) *Philosophy of Mind*. 2nd ed. Boulder, Colo.: Westview.

- Koch, Christof. (2004) *The Quest for Consciousness: A Neurobiological Approach*. Englewood: Roberts & Co. Publishers.
- Koeppsell, David R., and Lawrence S. Moss, eds. (2003) *John Searle's Ideas about Social Reality: Extensions, Criticisms, and Reconstructions*. Oxford: Blackwell.
- Kripke, Saul. A. (1971) "Naming and Necessity." In Donald Davidson and Gilbert Harman, eds. *Semantics of Natural Language*. Dordrecht, The Netherlands: Reidel, pp. 253–355; 763–769.
- Leibniz, Gottfried W. (2007) *Leibniz: The Monadology and other Philosophical Writings*. Translated by Robert Latta, Kessinger.
- Lepore, Ernest. (1995) "L'olismo della credenza e l'olismo del significato: John Searle su 'rete' e 'sfondo'" ("The Holism of Belief and the Holism of Meaning: John Searle on 'Network' and 'Background'"), *Rivista di filosofia*, 56:1, pp. 55–80.
- , and Robert Van Gulick (1991) *John Searle and His Critics*. Oxford: Blackwell.
- Levine, Joseph. (1997) "On Leaving Out What is Like." In Martin Davies and Glyn W. Humphreys, eds. *Consciousness: Psychological and Philosophical Essays*. Oxford: Blackwell, 1993, pp. 121–136; reprinted in Block, Flanagan, and Güzelçere, *The Nature of Consciousness*, pp. 543–555.
- . (2002) "Materialism and Qualia: The Explanatory Gap," *Pacific Philosophical Quarterly*, 64, 1983, pp. 354–361; reprinted in Chalmers, *Philosophy of Mind*, pp. 354–361.
- Liotti, Giovanni, ed. (2000) *Le discontinuità della coscienza: Etiologia, diagnosi e psicoterapia dei disturbi dissociativi (The discontinuities of Consciousness: Etiology, Diagnosis, and Psychoterapy of Dissociative Illnesses)*. 2nd ed. Milano: Franco Angeli.
- . (2005) *La dimensione interpersonale della coscienza (The Interpersonal Dimension of Consciousness)*. 5th ed. Milano: Carocci.
- Llinàs, Rodolfo. (2001) *The I of the Vortex*. Cambridge, Mass.: MIT Press.
- , and Denis Parè. (1996) "The Brain as a Closed System Modulated by the Senses." In Llinàs and Churchland, *The Mind-Brain Continuum*, pp. 1–18.
- , and Patricia Smith Churchland, eds. (1996) *The Mind-Brain Continuum: Sensory Processes*. Cambridge, Mass.: MIT Press.
- Locke, John. (1975) *An Essay Concerning Human Understanding*. Edited by P. H. Nidditch. Oxford: Clarendon Press.
- McCann, Edwin. (1995) "History: Philosophy of Mind in the Seventeenth and Eighteenth Century." In Guttenplan, *A Companion to the Philosophy of Mind*, pp. 338–347.
- . (1999) "Locke on Identity: Matter, Life, and Consciousness." In Atherton, *The Empiricists: Critical Essays on Locke, Berkeley, and Hume*, pp. 63–88.
- McDowell, John. (1991) "Intentionality De Re." In Lepore and Van Gulick, *John Searle and His Critics*, pp. 215–225.
- . (1994) *Mind and World*. Cambridge, Mass.: Harvard University Press.
- McGinn, Colin. (1997) "Can We Solve the Mind-Body Problem?" In Block, Flanagan, and Güzelçere, *The Nature of Consciousness*, pp. 529–542; originally published in *Mind*, 98:391 (1989), pp. 349–366.
- . (1999) "Can We Ever Understand Consciousness?" *New York Review of Books*, 46:10, pp. 44–48.
- Nagel, Thomas. (1981) "What Is It Like to Be a Bat?" In Hofstadter and Dennett, *The Mind's I*, pp. 393–403; originally published in *Philosophical Review*, 83 (1974), pp. 435–450.

- Nannini, Sandro. (2002) *L'anima e il corpo (The Soul and the Body)*. Roma and Bari, Italy: Laterza.
- . (2004) "Mental Causation and Intentionality in a Mind Naturalising Theory." In Peruzzi, *Mind and Causality*, pp. 69–96.
- Nida Rümelin, Martina. (2002) "Causal Reduction, Ontological Reduction, and First-Person Ontology: Notes on Searle's View about Consciousness." In Grewendorf and Meggle, *Speech Acts, Mind, and Social Reality*, pp. 205–221.
- Odegard, Douglas. (1970) "Locke and Mind–Body Dualism," *Philosophy*, 45, pp. 87–105.
- Parvizi, Josef, and Antonio Damasio. (2001) "Consciousness and the Brainstem," *Cognition*, 79, pp. 135–160.
- Paternoster, Alfredo. (2002) *Introduzione alla filosofia della mente (Introduction to the Philosophy of the Mind)*. Roma and Bari, Italy: Laterza.
- Peruzzi, Alberto, ed. (2004a) *Mind and Causality*. Amsterdam: Benjamins.
- . (2004b) "Causality in the Texture of the Mind." In Peruzzi, *Mind and Causality*, pp. 199–228.
- Piaget, Jean. (1972) *The Principles of Genetic Epistemology*. Translated by Wolfe Mays. New York: Basic Books.
- . (1975) *Biology and Knowledge: An Essay on the Relations between Organic Regulations and Cognitive Processes*. Translated by Beatrix Walsh. Chicago: University of Chicago Press.
- Pinker, Steven. (1994) *The Language Instinct*. Cambridge, Mass.: MIT Press.
- Popper, Karl R. (1994) *Knowledge and the Body-Mind Problem*. London, Routledge.
- , and John Eccles. (1977) *The Self and Its Brain: An Argument for Interactionism*, Berlin, London: Springer.
- Postman, Leo, Jerome Bruner, and Richard Walk. (1951) "The Perception of Error," *British Journal of Psychology*, 42, pp. 1–10.
- Preston, John, and Mark Bishop, eds. (2002) *Views into the Chinese Room*. Oxford: Oxford University Press.
- Putnam, Hilary. (1975a) *Mind, Language, and Reality*. *Philosophical Papers*, vol. 2. Cambridge, UK: Cambridge University Press.
- . (1975b) "The Meaning of 'Meaning'." In Putnam, *Mind, Language, and Reality*, pp. 215–271.
- . (1975c) "The Nature of Mental States." In Putnam, *Mind, Language, and Reality*, pp. 429–440; originally published as "Psychological Predicates." In William H. Capitan and David D. Merrill, eds. *Art, Mind, and Religion*. Pittsburgh, Penn.: University of Pittsburgh Press, 1967, pp. 37–48.
- . (1981) *Reason, Truth and History*. Cambridge, UK: Cambridge University Press.
- . (1988) *Representation and Reality*. Cambridge, Mass.: MIT Press.
- . (1996) "Introduction." In Andrew Pessin and Sanford Goldberg, eds. *The Twin Earth Chronicles*. New York: M. E. Scarpe, pp. xi–xxii.
- . (2000) *The Threefold Cord: Mind, Body, and World*. New York: Columbia University Press.
- Revonsuo, Antti. (1999) "Binding and the Phenomenal Unity of Consciousness," *Consciousness and Cognition*, 8, pp. 173–185.
- Rizzolatti, Giacomo, and Corrado Sinigaglia. (2006) *So quel che fai: Il cervello che agisce e i neuroni specchio (I Know What You Do: The Acting Brain and Mirror Neurons)*. Milano: Cortina.

- Rorty, Richard. (1970) "Incorrigibility as the Mark of the Mental," *The Journal of Philosophy*, 67:12, pp. 399–424.
- . (1979) *Philosophy and the Mirror of Nature*. Princeton: Princeton University Press.
- . (1980) "Searle and the Special Powers of the Brain," Commentary on Searle, "Minds, Brains, and Programs," *Behavioral and Brain Sciences*, 3:3, pp. 445–446.
- . (1982) "Contemporary Philosophy of Mind," *Synthese*, 53, pp. 323–348.
- . (1984) "What Is It About? Review of John Searle's *Intentionality*," *London Review of Books*, May 17–June 6, pp. 3–4.
- Rosciglione, Claudia. (2005) *Homo natura: Autoregolazione e caos nel pensiero di Nietzsche (Homo Natura: Self-Regulation and Chaos in Nietzsche's Thought)*. Pisa, Italy: ETS.
- Rosenthal, David. (1993) "Multiple Drafts Model and Higher-Order Thoughts," *Philosophy and Phenomenological Research*, 53:4, pp. 911–918.
- Rudder Baker, Lynne. (1993) "Metaphysics and Mental Causation." In Heil and Mele, *Mental Causation*, pp. 75–95.
- Searle, John R. (1969) *Speech Acts: An Essay in the Philosophy of Language*. Cambridge, UK: Cambridge University Press.
- . (1979a) *Expression and Meaning: Studies in the Theory of Speech Acts*. Cambridge, UK: Cambridge University Press.
- . (1979b) "A Taxonomy of Illocutionary acts." In Searle, *Expression and Meaning*, pp. 1–29; originally published in Keith Gunderson, ed. *Minnesota Studies in the Philosophy of Science*, vol. 7, *Language, Mind, and Knowledge*. Minneapolis, Minn.: University of Minnesota Press; reprinted in Searle, *Expression and Meaning*, pp. 1–29.
- . (1979c) "Literal Meaning." In Searle, *Expression and Meaning*, pp. 117–136; originally published in *Erkenntnis*, 13:1 (July 1978), pp. 207–224.
- . (1980) "The Background of Meaning." In John R. Searle, Ferenc Kiefer, and Manfred Bierwitsch, eds. *Speech Acts Theory and Pragmatics*. Dordrecht, The Netherlands: Reidel.
- . (1981) "Minds, Brains and Programs." In Hofstadter and Dennett, *The Mind's I*, pp. 353–372; originally published in *Behavioral and Brain Sciences*, 3:3 (1980), pp. 417–424.
- . (1983) *Intentionality: An Essay in the Philosophy of Mind*. Cambridge, UK: Cambridge University Press.
- . (1984) *Minds, Brains, and Science*. Cambridge, Mass.: Harvard University Press.
- . (1991a) "Response to Habermas." In Lepore and Van Gulick, *John Searle and His Critics*, pp. 89–96.
- . (1991b) "Response to Freeman and Skarda." In Lepore and Van Gulick, *John Searle and His Critics*, pp. 142–144.
- . (1991c) "Response to Armstrong." In Lepore and Van Gulick, *John Searle and His Critics*, pp. 181–185.
- . (1991d) "Response to Zemach." In Lepore and Van Gulick, *John Searle and His Critics*, pp. 188–192.
- . (1991e) "Response: Reference and Intentionality." In Lepore and Van Gulick, *John Searle and His Critics*, pp. 227–230.
- . (1991f) "Response to Burge." In Lepore and Van Gulick, *John Searle and His Critics*, pp. 231–237.

- . (1991g) "Response to McDowell." In Lepore and Van Gulick, *John Searle and His Critics*, pp. 237–241.
- . (1991h) "Response to Stroud." In Lepore and Van Gulick, *John Searle and His Critics*, pp. 288–295.
- . (1992) *The Rediscovery of the Mind*. Cambridge, Mass.: MIT Press.
- . (1995a) "Consciousness, the Brain and the Connection Principle: A Reply." *Philosophy and Phenomenological Research*, 55:1, pp. 217–232.
- . (1995b) *The Construction of Social Reality*. Hardsworth, UK: Penguin Books.
- . (1997) *The Mystery of Consciousness*. London: Granta Books.
- . (1998) *Mind, Language and Society: Philosophy in the Real World*. New York: Basic Books.
- . (2001) *Rationality in Action*. Cambridge, Mass.: MIT Press.
- . (2002a) *Consciousness and Language*. Cambridge, UK: Cambridge University Press.
- . (2002b) "How to Study Consciousness Scientifically." In Searle, *Consciousness and Language*, pp. 18–35; originally published in *Philosophical Transactions of the Royal Society*, ser. B. 353:1377 (29 November 1998), pp. 1935–1942.
- . (2002c) "Consciousness." In Searle, *Consciousness and Language*, pp. 36–60; originally published in *Annual Review of Neuroscience*, 23 (2000), pp. 557–578; reprinted in Searle, *Consciousness and Language*, pp. 36–60.
- . (2002d) "Collective Intentions and Actions." In Searle, *Consciousness and Language*, pp. 90–105; originally published in Philip R. Cohen, Jerry Morgan, and Martha. E. Pollack, eds. *Intentions in Communication*. Cambridge, Mass.: MIT Press 1990.
- . (2002e) "Why I am not a Property Dualist," *Journal of Consciousness Studies*, 9:12, pp. 57–64.
- . (2002f) "Animal Minds." In Searle, *Consciousness and Language*, pp. 61–76; originally published in *Midwest Studies in Philosophy*, Vol. 19, 1994, pp. 206–219.
- . (2002g) "Analytic Philosophy and Mental Phenomena." In Searle, *Consciousness and Language*, pp. 203–225; originally published in Peter A. French, Theodore E. Uehling, Jr., and Howard K. Wettstein, eds. *Midwest Studies in Philosophy*, 6, 1981, Minneapolis: Minnesota University Press 1981.
- . (2002h) "End of the Revolution." Review of Noam Chomsky, *New Horizons in the Study of Language and Mind*. *The New York Review of Books*, 28 February, pp. 33–36.
- . (2004) *Mind: A Brief Introduction*. Oxford: Oxford University Press.
- . (2005) "The Quest for Consciousness." Review of Christof Koch, *The Quest for Consciousness*. *The New York Review of Books* (13 January).
- . 2006, "What is Language: Some Preliminary Remarks." In Günther Abel, ed. *Kreativität (Kolloquienbeiträge)*. XX. *Deutscher Kongreß für Philosophie*. Berlin: Felix Meiner Verlag, pp. 1223–1248.
- , and Walter Freeman. (1998) "Do We Understand Consciousness?" *Journal of Consciousness Studies*, 5:5–6, pp. 718–733.
- Sellars, Wilfrid. (1963) "Philosophy and the Scientific Image of Man." In *Science, Perception, and Reality*. London: Routledge & Kegan Paul, pp. 1–40.
- Shear, Jonathan, ed. (1995) *Explaining Consciousness: The Hard Problem*. Cambridge, Mass.: MIT Press.
- Shoemaker, Sydney. (1993) "Lovely and Suspect Ideas," *Philosophy and Phenomenological Research*, 53:4, pp. 905–910.

- Singer, Wolfe. (1996) "Neuronal Synchronization: A Solution to the Binding Problem?" In Llinàs and Churchland, *The Mind-Brain Continuum*, pp. 101–130.
- Smith, Barry, ed. (2003) *John Searle*. Cambridge, UK: Cambridge University Press.
- Sparti, Davide. (1996) *Soggetti al tempo: Identità personale tra analisi filosofica e costruzione sociale (Subjects to Time: Personal Identity between Philosophical Analysis and Social Construction)*. Milano: Feltrinelli.
- . (2001) *Identità e coscienza (Identity and Consciousness)*. Bologna: Il Mulino.
- Stoutland, Frederick. (1994) "Searle's Consciousness." Review of John R. Searle's *The rediscovery of the Mind*. In *Philosophical Books*, 35:4, pp. 245–254.
- Sutherland, Stuart. (1989) *The International Dictionary of Psychology*. New York: Continuum.
- Tagliagambe, Silvano. (1980) *La mediazione linguistica: rapporto pensiero–linguaggio da Leibniz ad Hegel (The Linguistic Mediation: Thought–Language Relationship from Leibniz to Hegel)*. Milano: Feltrinelli.
- Thompson, David L. (1986) "Intentionality and Causality in John Searle," *Canadian Journal of Philosophy*, 16, pp. 83–97.
- Tomida, Yasuhiko. (2001) *Inquires into Locke's Theory of Ideas*. Hildersheim, Germany: Georg Olms Verlag.
- Tononi, Giulio. (2003a) *Galileo e il fotodiodo (Galileo and the Photodiode)*. Roma and Bari, Italy: Laterza.
- . (2003b) "Consciousness Differentiated and Integrated." In Cleeremans, *The Unity of Consciousness*, pp. 253–265.
- Tye, Michael. (1993) "Reflections on Dennett and Consciousness," *Philosophy and Phenomenological Research*, 53:4, pp. 893–898.
- . (1995) *Ten Problems of Consciousness: A Representational Theory of the Phenomenal Mind*. Cambridge, Mass.: MIT Press.
- Varela, Francisco, Evan Thompson, and Eleanor Rosch. (1991) *The Embodied Mind: Cognitive Science and Human Experience*, Cambridge, Mass.: MIT Press.
- Viano, Carlo Augusto. (1960) *John Locke: Dal razionalismo all'illuminismo (John Locke: From Rationalism to Enlightenment)*. Torino, Italy: Einaudi.
- Viskovatoff, Alex. (2001) "Searle's Background: Comments on Runde and Faulkner," *Journal of Economic Methodology*, 9:1, pp. 65–80.
- Walter, Sven, and Heinz Dieter Heckmann, eds. (2003) *Physicalism and Mental Causation*. Exeter: Imprint Academic.
- Weiskrantz, Larry. (1986) *Blindsight*. Oxford: Oxford University Press.
- . (1988) "Some Contributions of Neurophysiology of Vision and Memory to the Problem of Consciousness." In Anthony J. Marcel and Edoardo Bisiach, *Consciousness in Contemporary Science*.
- Yolton, John. (1991) *Locke and French Materialism*. Oxford: Clarendon.
- Zanet, Giancarlo. (2007) *Le radici del naturalismo: W. V. Quine tra eredità empirista e pragmatismo (The Roots of Naturalism: W. V. Quine between Empiristic Inheritance and Pragmatism)*. Macerata, Italy: Quodlibet.
- Zeki, Semir. (2003) "The Disunity of Consciousness," *Trends in Cognitive Sciences*, 7:5, pp. 214–218.

ABOUT THE AUTHOR

Giuseppe Vicari graduated in Philosophy in 2004 and completed the PhD in Philosophy in 2007 at the University of Palermo, Italy, under the supervision of Francesca Di Lorenzo Ajello. His dissertations addressed the problem of realism in John Locke's *Essay Concerning Human Understanding* and John R. Searle's philosophy of mind in relation to contemporary debate in philosophy and neurosciences. His research interests are in History of Philosophy (with special emphasis on the History of Modern and Contemporary Philosophy) and Philosophy of Mind.

Vicari was Visiting Scholar with the title of Visiting Student Researcher at the Department of Philosophy of the University of California at Berkeley during the Spring Semester 2007, sponsored by Searle. He also spent other research time at Berkeley during the Spring Semester 2005 and September 2004 while working on his doctorate.

Currently, Vicari is a member of the research group in History of Philosophy coordinated by Francesca Di Lorenzo Ajello (*Dipartimento di Filosofia, Storia e Critica dei Saperi*, University of Palermo, Italy), a member of the *European Society for Analytic Philosophy*, *Società Italiana di Filosofia Analitica* (Italian Society of Analytic Philosophy), the *Società Filosofica Italiana* (Italian Philosophical Society), and the Berkeley Social Ontology Group.

Vicari has authored papers in the *Proceedings of the XVI National Conference of PhD in Philosophy* (Reggio Emilia, Italy 2005) on John Searle's philosophy of mind and as part of the *Proceedings of the Rationality, Truth, and Mind* series of lectures and seminars (Palermo, Italy, 2006), edited by Francesca Di Lorenzo Ajello, on the problem of the self. He has also translated John Searle's paper "What is Language," into Italian for the *Proceedings of Rationality, Truth, and Mind*.

INDEX

- abilities, 19, 24, 112, 113, 129, 132, 150
 - Background a., 124–128, 156, 162
 - biological a., 128, 162
 - cognitive/mental a., 36, 85, 109,
 - 111, 124, 136, 137, 139, 146, 156, 159
 - conceptual a., 157
 - (pre)linguistic/semantic a., 108, 124
 - (non)(pre)intentional a., 109, 132
- aboutness, 18
- action(s), 2, 6, 7, 13, 15, 26–29, 31, 35,
 - 41, 56, 69, 76, 77, 98, 105, 112, 114, 125, 127, 131, 132, 145, 150–153, 156, 157, 162
- box and arrow models of action, 125
- causal a., 6, 64, 107
- collective a., 102, 120, 122, 123, 128
- cooperative a., 122, 124
- future a., 30, 151
- gap between a. and its antecedents,
 - 152, 154
- goal-directed a., 127
- intentional a., 57, 98, 104, 107, 108,
 - 116, 125, 145, 153
- past a. 28
- possible a., 127, 128
- primitive a., 147
- rational a., 153, 154
- a. schemes, 109, 125
- visually encoded a. patterns, 128
- agen(cies)(cy)(t)(s), 2, 22, 26, 27, 45,
 - 77, 84, 102, 111, 120, 122, 141, 153, 155, 159
- animal a., 153
- a. capable of organizing provisional
 - coalitions, 74
- cognitive a., 64, 76
- conscious a., 45, 122
- cooperative a., 122
- executive a., 142
- human a., 142
- intelligent a., 28
- potential a., 128
- speaker-a., 118, 120
- agnosticism, 31
- Alexander, Samuel, 51
- Alexander's dictum, 13
- alien hand syndrome, 81
- alien thoughts, 81
- amygdala, 150
- angiotensin, 53
- animals, 11, 12, 49, 76, 122, 157
 - human a., 101, 111, 121
 - (non)(pre)linguistic a., 110, 111
 - pack a., 122
 - social a., 124
- Anscombe, Elizabeth, 104
- antiCartesian view, 27
- antireductionis(m)(ts), 3, 9, 12, 14, 15,
 - 17, 20, 22–24, 35, 43, 48, 57, 58, 64, 65, 67, 69, 70, 87, 92, 93, 97, 98, 139, 159, 161
- antisubstantialist view, 25, 27, 33
- Apel, Karl O., 115, 116
- appearance(s), 62, 80, 147
 - a-reality distinction, 61, 84
 - subjective a., 60, 61
- Aristotle, 19, 119
- artificial intelligence (A.I.), 18, 108
 - strong/weak A. I., 20, 21
- aspectual shape, 104, 105, 125, 126
- assertives, 102
- Austin, John L., 104
- Awareness, 39, 40, 135, 161
- Background, 6, 84, 102, 109–121, 123–128, 154, 156. *See also* figure-background structure, Network, self-referentiality
 - pragmatic B., 105
 - preintentional B., 128, 132, 162
 - subjective B., 80
- bat physiology, 12
- Bechtel, William, 48
- Bedeutung* (reference), 105. *See also* Sinn (sense)
- behavior(ists)(m)(s) (also behaviour),
 - 5, 6, 15–23, 36, 37, 44, 47, 50, 53, 56, 65, 86, 89, 109, 113, 114, 129, 132, 142, 145, 148, 155–157, 159, 160, 162
- brain-b., 56, 141
- causal-b., 22, 51

- behavior(ists)(m)(s) (also behaviour), *con't.*
 collective/cooperative b., 120, 121,
 122, 125, 128
 complex b., 56, 75
 conscious b., 57
 b. creativity and flexibility, 57,
 132, 155
 explanation of b., 94
 external b., 55, 56
 goal-directed, 148
 intelligent b., 97, 112
 linguistic-conversational b., 122
 molecular b., 58, 69
 b. outputs, 18–20, 22, 56
 belief(s), 6, 15, 20, 29, 36, 42, 51, 65,
 67, 75, 78, 82, 83, 91, 103–
 106, 108, 110, 111, 121, 143,
 152, 158
 dispositional b., 144
 mistaken b., 104
 Bennett, Jonathan, 107
 binding, conceptual, 126
 b. device, 46, 131
 semantic b., 127
 binocular rivalry, 134
 biological processes, 2, 27, 49, 50, 58,
 151, 157, 160, 162. *See also*
 neurobiological processes
 blindsight, 16, 76, 81, 133, 134
 Block, Ned, 20, 22, 41, 42, 85, 86, 89
 B.'s "access consciousness," 85
 bootstrapping effect, 108, 117, 124, 156
 box and arrow models of action, 125.
 See also action(s)
 brain(s), 4, 5, 7, 10–12, 18, 19, 53, 54,
 74, 114, 138
 b. anatomy/architecture, 95, 128,
 131, 160
 animal b., 50, 155, 157
 b. architecture, 95, 128
 b.-behavior functions, 56
 biologically embodied system
 features of b., 55, 128, 141,
 146, 150
 Cartesian Theater in the b., 81,
 147, 160
 b. categorization activity, 46, 135
 b.-mind relationship, 49, 56, 58,
 70, 134
 b. celebrity, 142
 computational speed of b., 112
 conscious b., 48, 54, 56, 92, 93, 99,
 116, 132–134, 137, 139, 142,
 149, 150
 nonconscious process of b., 151
 b. damage, 36, 111, 133, 137
 b. (electro)chemistry, 36, 63
 b. events vs. mental events, 37
 fame in the brain (cerebral fame), 76
 b. functions/processes, 5, 10, 44, 64,
 94, 145
 functional disorders of b., 137
 learning molded by b., 143
 measuring b. activity, 35
 b. memory systems, 46
 mirror neurons in b., 129
 b. motor areas, 127
 movie-in-the-b., 147
 neurobiological processes in b., 47,
 89, 136
 b. as organic machine, 21
 b. organization, 155
 pain caused by b., 61
 self-monitoring b. devices, 162
 semantic abilities of b., 108, 126
 b. vs. soul, 140
 split b., 76, 81, 134, 137, 142
 b.stem, 13, 139, 141, 152
 b.(s) in (a) vat(s), 115–117, 120, 128
 virtual machine in b., 75
 b. wiring diagram, 19
 Brandom, Robert B., 110
 bridge laws, psychophysical, 17, 22,
 86, 95, 160
 Broad, Charles D., 51
 building block approach to conscious-
 ness, 131, 133, 138. *See also*
 unified field model
 Burge, Tyler, 16, 65, 67, 156
 care, 2
 self-c., 28, 30

- Cartesian Theater, 73–76, 78, 79, 81, 95,
138, 142, 145, 147, 160, 162
- Cartesianism, 13, 83
- categorical mistake, 26, 36, 75, 106
- categor(ies)(ization)(y) 33, 46, 47,
124–127
- brain c., 46
 - Cartesian c., 34, 71
 - conceptual c., 146
 - c. of empirical reality, 45
 - linguistic c., 126
 - c. mapping, 126
 - metaphysical c., 31
 - c. mistake, 26, 36, 75, 106
 - mutually exclusive c., 41, 42
 - ontological c., 4, 73
 - c. of “other people,” 116
 - perceptual c., 109, 135, 136, 139
 - memory re-categorization of c.,
136, 139
 - self-c., 135
 - self-identifying c., 124
 - value-c. systems, 46, 135, 139
- causal closure, 3, 23, 24, 47, 62, 64–
70, 161
- causal exclusion dilemma, 22, 55, 57,
62, 68, 96, 101
- causal relations, 3, 13, 17, 19, 22, 37,
51, 55, 56, 97, 98, 107, 119,
145, 158
- causal-behavioral analysis, 22
- causality, 37, 45, 98, 106, 107, 114, 152
- causally closed physical world, 15, 96
- causation, 15, 17, 38, 52, 60, 68
- billiard ball c., 154
 - direction of c., 107
 - bottom-up c., 37, 52, 57
 - b.-u. no time gap c., 54, 55,
69, 73, 95, 155, 161
 - top-down (downward) c., 69, 70
- immanent c., 154
- intentional c., 108
- mental c., 1, 3, 6, 9, 13, 15–17, 20,
22, 24, 33, 35, 39, 47, 48, 55–
57, 62, 64–68
- nomological theory of c., 54
- physical c., 152
 - simultaneous c., 37
 - standard theory of c., 34, 36, 37, 48,
- centering, 2
- cerebellum, 10
- cerebral fame (fame in the brain), 76
- certainty, 31
- epistemic c., 83, 85, 95
- Chalmers, David J., 1, 2, 5, 6, 10, 13,
14, 17, 21, 24, 35, 44, 52, 64,
68, 73, 85–97, 101, 142–145,
159–161
- Changeux, Jean Pierre, 134
- Chinese nation argument, 20, 89
- Chinese room argument, 20, 21, 33
- Chisholm, Roderick M., 107, 153
- Chomsky, Noam, 18, 117
- Churchland, Patricia Smith, 14, 134
- Churchland, Paul, 14
- Clark, Andy, 142, 144, 145
- cognition, 26, 30, 96, 125, 127, 143
- Cognition*, 159
- cognitive capacities, 112, 129
- cognitive science, 1, 4, 18, 64, 65, 79–
81, 85, 86, 142
- “Collective Intentions and Actions”
(Searle), 120
- combinatorial explosion, 112
- commissives, 102
- computational model, 78, 112, 113
- computationalists, 18
- computerized axial tomography
(CAT), 10
- conscious life, organism, 6, 10, 11, 36,
46, 53, 102, 127, 131
- consciousness/mind, *passim*
- building block model of c., 135
 - ontology of c./m., 2, 3, 5, 6, 22, 24,
33, 38, 39, 46, 47, 49, 56–58,
68, 95, 99, 114, 152, 158, 159;
see also ontology
 - c. states, 1, 4, 16, 20, 35, 41, 43,
45–48, 51, 53, 59, 74, 83, 94,
131, 132, 137, 138
 - A c. states, 42
 - phenomenal (*P*) c., 41, 42

- consequences, 28, 30, 49, 64, 128, 133, 155, 157, 158, 162
 causal c., 59
 logical c., 34
 metaphysical c., 59
 constitutive rules, 121, 123, 124
 content determination, 105, 118–120, 132
 context-dependence, 109, 110
 contextualism, 25
The Construction of Social Reality
 (Searle), 120
 corpus callosum, 137
 Crane, Tim, 43, 44, 82, 160
 creativity, 57, 146, 151, 152, 155, 162
 Crick, Francis, 85, 133, 134
 culture, 71, 74, 76, 77, 117, 124, 141
- Damasio, Antonio, 6, 7, 10, 28, 58, 99, 131, 132, 139, 146–152, 157, 158, 162
- Darwinism, 64
 neuronal D., 135
- data, 11, 76, 112, 113, 127–129, 137, 141
- Davidson, Donald, 22, 24, 35, 52, 63, 87, 107, 110, 157
- decision-making, 140
- declarations, 102
- Deep Blue, 112
- Dennett, Daniel C., 1, 2, 5, 6, 14, 19, 21, 24, 33, 43, 48, 65, 73–83, 85, 95–98, 101, 112, 138, 141, 142, 144, 145, 155, 157, 159–161
- deontic powers, 122, 123
- Descartes, René, 13, 15, 25, 26, 30, 33, 45, 54, 83, 91, 95, 106, 117, 138, 140, 141
- description(s), 12, 13, 18, 19, 20, 21, 35, 50, 55, 58, 66, 68, 92, 93
 causal d., 69
 contradiction in d., 89
 levels of d., 52, 54, 70, 91
 mental d., 119
 misleading d., 84
 d.-organization, 57
 re-d., 144
 third-person d., 58
- design, 25, 30, 75, 77, 124
- desires, 6, 51, 75, 103–105, 107, 108, 110, 118, 121, 143, 152
- desubstantialization of mind, 1, 3–6, 9, 17, 19, 21, 22, 24, 25, 33, 39, 49, 73, 80, 85, 95–97, 101, 132, 140, 147, 159, 162
- Di Francesco, Michele, 48, 81, 82, 144
- Di Lorenzo Ajello, Francesca, 6, 80, 98, 116
- digestion, 27, 49, 50, 51, 58, 93
- directives, 102
- dissociative syndromes, 81
- divine mediation, Melebrache's, 15
- DNA, 59, 77
- double aspect theory, Nagel's, 13
- duali(m)(ts), 22, 28–32, 35, 51, 55, 65, 92–94, 96, 98, 159, 161, 163
 Cartesian d. (pluralism), 5, 13, 26, 75, 95
 conceptual d., 4–6, 9, 33, 34, 36, 39, 48, 49, 54, 57, 68, 70, 73, 95–97, 99, 101, 152, 155, 160, 161
 monism-d. dilemma, 3, 4, 9, 18, 24, 25, 33, 38, 39, 54, 68, 73, 95, 101, 160
 neod., 15, 23, 24, 64, 159
 Chalmer's neod., 85, 91, 95
 d. ontology, 95
 Popper's d., 13
 property d., 1, 2, 13, 14, 21, 34, 58, 64, 85, 88–92, 95, 98, 161
 substance d., 34, 45, 91
- dynamic core, 126, 131, 135–137, 139
- Eccles, Sir John, 14, 48
 The Self and Its Brain (with Popper), 14
- echolocalization, 12, 58
- Edelman, Gerald M., 7, 45–47, 58, 76, 85, 99, 108, 126, 131, 132, 135–139, 146, 162
 The Remembered Present, 146
 A Universe of Consciousness (with Tononi), 135

- eliminativis(m)(ts), 16, 17, 80, 95, 96–98, 139
- embodiment, 6, 19, 24, 33, 141, 146, 147, 158, 160
 - e. of agent, 102
 - e. of consciousness, 132, 138, 146, 152, 157–159, 162
 - criterion of e. for definition of self, 142
 - cultural e., 117
 - e. of intentional content, 106
 - e. of mind, 57, 101, 130, 136, 146, 150, 155, 156
 - e. of Network, 125
 - e. of perception, 126
- emergentism, 51
- emotion(s), 1, 10, 35, 36, 41, 51, 74, 128, 150, 155
- epilepsy, 16, 137, 148
- epiphenomenalis(m)(t), 3, 16, 17, 35, 64, 65, 67, 68, 70, 94, 98, 99, 146, 161
- epistemologist, 13
- Essay Concerning Human Understanding* (Locke), 3, 24, 25
- evolutionary advantage, 7, 57, 77, 131, 151, 152
- exclusion principle, 4, 34, 36, 37, 63, 68
- expectation, 126, 129, 141
- experience(s), 42, 56, 80, 81, 96, 118, 126, 162
 - e. of agent, 45
 - atoms of e., single, 135
 - auditory e., 35
 - conscious e., 17, 40, 41, 43, 46, 47, 55, 74, 75, 78, 82, 86, 88, 89, 90, 91, 93–95, 104, 113, 116, 125, 126, 133, 134, 136–138, 160
 - cultural e., 91
 - existence of subject with e., 101, 131
 - objective aspects of e., 13
 - perceptual/sensory/phenomenal e., 15, 42, 58, 104–106, 107, 135, 155–157
 - taste e., 40, 41, 118
 - visual e., 79, 84, 105, 106, 108, 109, 119, 134, 136
 - personal e., 145
 - phenomenal e., 90
 - primordial e., 108
 - repeated e., 141
 - sensible e., 27
 - sequence of e. that includes a decision, 154
 - subject of e., 77
 - subjective character of e., 12, 36, 60–62, 78, 85, 93, 129
 - unity of e., 81
- explanation(s), 154
 - alternative e., 113
 - e. of behavior, 65, 70, 85, 86, 94
 - bottom-up no time gap e., 95
 - bottom-up micro-macro e., 53, 54
 - e. how brain produces a self, 10
 - causal e., 17, 37, 54, 62, 63, 64, 69, 152, 153
 - e. in classic mechanics, 154
 - e. of cognitive phenomena, 85
 - e. how conscious feelings fit into worldview, 53
 - intentional e., 66
 - left-right macro-macro e., 53
 - mental e., 65–67
 - e. in ontological terms, 66
 - physical e., 16, 66, 92
 - e. of psychophysical relation, form of, 5
 - rational e., 152–154
 - e. of rational action, 151, 152, 154
 - reductive e., 87, 88
 - scientific e., 79, 97
- explanatory gap, 9–11, 13, 14, 35, 36, 91, 154
- Expression and Meaning* (Searle), 102
- expressives, 102
- externalism, 117, 118, 120
- facts, 57, 61, 85, 149, 150
 - biological f., 87, 161
 - f. about consciousness, 89
 - cultural f., 128
 - f. independent of human beings, 111
 - institutional f., 121, 123–125

- facts, *con't.*
 knowable/understandable f., 97
 linguistic elements of f., 125
 low-level f., 88
 objective f., 12, 122
 physical f., 34, 85, 87–90, 160, 161
 physiological f., 58
 positive f., 88, 89
- facult(ies)(y), 2, 19, 27
 f. of body/mind, 28
 digestive f., 27
 elective f., 27
 expulsive f., 27
 intellectual f., 27
 f. of introspection, 83
 f. of the mind, 28
 motive f., 27
 f. of thinking, 31
 f. of understanding and will, 30
- falsification of concepts, 75
- Feyerabend, Paul K., 14
- figure-background structure, 46, 125
- fit(ness), 14, 143, 154
 creative f., 113
 direction of f., 102–105, 107, 111, 118
 mind-to-world direction of f., 105, 107
 inclusive f., 53
- Fodor, Jerry, 15, 16, 18, 24, 65, 67, 78, 79
- force fields, 9
- frame problem, 113
- Freeman, Walter, 36, 114
- Frege, Gottlob, 105, 119
- functional magnetic resonance (fMRI), 10
- functionalism, 1–5, 9, 17–24, 33, 39, 48, 64, 65, 85, 86, 95, 101, 112. *See also* mind: desubstantialization of; reduction(ism)
 antireductionist f., 22
 computational f., 1
 homuncular f., 112
 Putnam's style f., 20
- generalization argument, 67
- genius, Descartes' evil, 117
- genotype, 54
- Gestalt psychologists, 46, 125
- Gestalt Theory, 109
- Gestalten* (semantically loaded figures), 125
- Ghost in the Machine, 44, 75, 140, 142, 154
- God, 11, 30, 35, 87, 88, 90, 93
- gravity, 37, 140, 141, 145
- grey matter, 11. *See also* brain, thinking
- Habermas, Jürgen, 116
- hippocampus, 150
- Hobbes, Thomas, 1, 33
- Hofstadter, Douglas, 21, 48
- holism, 3, 6, 101, 103–129 *passim*, 131, 132, 138, 139, 146, 152, 156, 158
- Holt, Lynn, 28
- homunculi, 108
- Hume, David, 45, 140, 141, 153
 neo-H. line of thought/view, 141
- Huxley, Thomas H., 35
- hypostatization, 36, 134, 152
- "I," 45, 132, 140, 152, 154, 155–157
- identity, 27, 28, 34, 46, 58, 91, 149, 150
 i. of biological processes, 2
 i. of brain/mind causal powers, 94
 i. of consciousness, 2
 i. dissociation, 81, 142
 i. dysfunctions, 142
 i. of mental states, 19, 46
 mind-body i., 65
 performance i., 20
 personal i., 28, 144
 i. of self, 2
 i. of states, 101, 115
 i, theor(ists)(y) (identitists), 16, 18, 48, 68, 91
 token i. theory, 17, 19, 22–24
 type i. theory, 3, 4, 22, 69, 115
 i. statements, 13
- illocutionary forces, 103
- Gage, Phyneas, 36

- image, manifest, 15
- imaging techniques, 10, 38
- incorrigibility, 43, 78, 82–84, 145
- indexicality, 119, 156, 157
- information, 16, 56, 58, 144
 - accurate i. about the world, 111
 - background i., 76
 - conscious i., 81
 - internal mental states vs. that stored
 - in external sources, 143
 - location in brain where i. comes
 - together, 78
 - peripheral modules, i. processed
 - by, 78
 - i. processing, 18, 127, 133, 145, 160
 - sensory i., 2, 127
 - partial sensory i., 133
 - stored i., 143, 145
- innatism, 117
- input-output patterns, 22
- institutions, 122
- intelligence, 50, 112, 113
 - artificial i. (A.I.), 18, 108
 - strong/weak A. I., 20, 21
- intention(s), 6, 35, 56, 67–69, 75, 98, 104,
 - 105, 107, 108, 110, 111, 118,
 - 121, 127, 129, 132, 152, 157
- intentional states, 6, 42, 75, 84, 98,
 - 103–115, 118, 121, 124, 132,
 - 156, 158, 162
- paradigmatic i. s., 105
- primitive i. s., 108, 162
- intentionality, 6, 18, 47, 49, 98,
 - adult i., 108
 - collective i., 116, 120–124, 128
 - conscious i., 126
 - formal structure of i., 102, 125, 126
 - holistic structure of i., 156
 - “of” of i., 43
 - prelinguistic i., 108, 111, 112, 125, 156
 - primitive forms of i., 158
 - Searle’s theory of i., 99, 101–129,
 - 131, 156, 162
- Intentionality* (Searle), 80, 112, 162
- interactionism, Cartesian, 62
- internalism, 117, 118, 120
- introspection, 11, 44, 80, 83, 84, 140
- Jackendoff, Ray, 78, 112
- Jackson, Frank, 12, 20, 22, 35 58, 64,
 - 65, 68
- James, William, 18, 137
- Kant, Immanuel, 45, 106
- Kasparov, Garry, 113
- Kim, Jaegwon, 2, 3, 5, 6, 16, 21–24,
 - 48, 52, 53, 55, 57, 62–64, 66–
 - 70, 82, 87, 96–98, 101, 146,
 - 159–161
- K.’s causal exclusion dilemma, 22,
 - 55, 57, 62, 68, 96, 101
- knowledge, 15, 25, 26, 31, 32, 64, 67,
 - 83, 136, 150
 - scientific k., 80
 - semantic k., 126
- Koch, Christof, 85, 133, 134
- Korsgaard, Christine M., 153
- Kripke, Saul A., 13, 82, 87, 90, 93
- language, 21, 42, 76, 85, 96, 102, 105,
 - 108, 116, 121, 125, 141, 143,
 - 156, 160
- brain damage in l. areas, 111
- capacity for l., 50
- dependence of mind on l., 110,
 - 111, 115
- l. game of explaining action, 154
- l. headquarters in brain, 160
- institutional nature of l., 124
- intentional l., 144
- natural l., 108, 109, 111
- ordinary l., 34, 91
- private l. argument, 74, 78, 80
- public l., 74, 80, 106, 111
- social conception of l., 117, 124
- thought, l.-dependence of, 125, 124
- Leibniz, Gottfried W., 1, 10, 11, 15,
 - 28, 31, 35
- levelism, 69
- Levine, Joseph, 9, 13, 35
- life, 2, 4, 5, 29, 47, 78, 149, 152, 162
 - biological l., 155

- life, *con't.*
 conscious l., 6, 10, 11, 36, 46, 53,
 102, 127, 151
 mental l., 43, 55, 81, 117
 nature, relations between l. and, 91
 physical world, relations between l.
 and, 91
 l. not supervient on physics, 161
 liquidity, 50, 51, 54, 58, 69, 93
 Llinàs, Rodolfo, 134
 location in space, 36
 Locke, John, 2, 3, 19, 159
 L.'s agnosticism, 31
 L.'s antisubstantialist turn, 33
 L. as a dualist, 28, 29, 32
 L. as a materialist, 28,
 L.'s philosophy of mind, 9, 25–31, 33
 L.'s skepticism, 31
 subjectivity, L.'s picture of, 32
 Essay Concerning Human Under-
 standing (Locke), 3, 24, 25
- Mach, Ernst, 19
 macrofeature(s), 4, 37, 54, 55, 73,
 160, 161
 macrostructural states or processes, 4
 Malebranche, Nicolas, 15
 mark test, 157
 Mass-energy points, 145
 materialis(m)(ts), 1–6, 12–14, 17, 19,
 28, 37, 87–89, 96, 155
 Cartesian m., 79
 consciousness, capability of m. to
 explain, 85, 88, 91, 95, 101
 dualism, relation between m. and,
 14, 18, 34, 36, 48, 96, 159
 eliminative m., 14, 17
 foundational m. theory, 18
 functionalism, m. part of, 24
 Kim's m., 5
 Locke's m., 28–31
 imm., 35
 mentalism vs. m., 35
 m. metaphysics, 65
 m. philosophical orthodoxy, 15, 18
 matter, 2, 5, 28, 35
 extended m., 32
 insensible m., 29
 realizing m., 27
 Searle's m., 48, 58
 senseless m., 31
 McCann, Edwin, 31
 McDowell, John, 17, 110, 115, 116, 157
 McGinn, Colin, 11, 12, 48, 159
 meme(s), 76, 77
 memor(ies)(y), 40, 42, 46, 55–57, 98,
 107, 112, 118, 132, 141, 145,
 146, 149, 150
 conceptual m., 135
 re-categorization function of m.,
 136, 139
 mental, the, 3, 5, 12, 35, 36, 50, 52, 53,
 70, 97, 114, 155, 159, 160,
 161. *See also* physical, the
 abilities, m., 111, 124
 linguistic m. abilities, 109
 primitive m. abilities, 156, 162
 body, m. substance interact with, 32
 categories, m. and physical are
 mutually independent and
 exclusive, 4, 73
 m. caus(ation)(es), 3, 6, 9, 13, 15–
 17, 20, 22–24, 33, 35, 39, 47,
 48, 55–57, 62, 64–71, 94, 96,
 97–99, 101–129, 131, 146,
 152, 156, 158, 160, 163
 m. content of person on earth, 118
 m. descriptions, 119
 distinction between m. and brain
 events, 37
 distinction between m. and physical,
 9, 10, 14, 15, 19, 23, 28, 34,
 35, 37, 48, 51, 52, 61, 70, 71,
 92, 93, 98, 101
 environment as m., 143
 exclusion principle of m. and
 physical, 68)
 explanations, m. and physical, 65,
 66
 m. external supports, 142
 m. images, 147
 m. life, 43, 55, 117

- mental, the, *con't.*
 conscious m. life, 81
 nonrepresentational m. capacities,
 84, 111
 m. organization, 145
 m. patterns of an object, 147, 148
 m. phenomena, 3, 9, 10, 12, 15, 17,
 18, 19, 21, 24, 29, 49, 55, 58,
 97, 110, 111, 115, 138, 155
 conscious m. p., 56
 physical, dependence of mental on, 67
 preintentional m. skills, 125, 156
 m. processes, 2, 19, 20, 22, 36,
 49, 97, 108, 128, 148, 159
 m. primitive processes, 157
 unconscious m. processes, 143
 representations, (un)conscious m.,
 102, 110, 113
 m. scenes, 132
 m. state(s), 1–3, 6, 12, 13, 15–17,
 19–22, 36, 46, 47, 49, 56, 66,
 68, 75, 80, 83, 86, 96, 97, 101,
 102, 105, 106, 109, 111, 114,
 119, 126, 127, 140, 145, 151,
 152, 159, 161
 brain state, m. state as, 58
 cognitive m. states, 30
 conscious m. states, 36, 54,
 56, 156
 first-person ontology of m.
 states, 155
 higher-level/order m. activities,
 78, 94
 holistic interdependence be-
 tween m. states, 125
 intentional m. states, 21
 internal m. states, 145
 network of m. states, 46, 127, 158
 neurophysiological states, m.
 states supervenient on, 51
 Pleistocene man's m. state, 115
 powers, m. states as, 27
 primitive m. states, 98, 108, 114
 process-like nature of m.
 states, 146
 propositional content, m. states
 with (thoughts), 42
 qualitative (phenomenal) m.
 state, 78
 relational view of m. states, 27
 subjective m. states, 4, 35, 44,
 85, 155
 (un)conscious m. states, 12, 20,
 36, 56
 microphysics, 4, 35, 91
 microstructural processes, 4
 mind, ontology of. *See* conscious-
 ness/mind: ontology of
 mind-body problem (interaction be-
 tween mind and body), 1, 2, 9–
 38, 43, 47–49, 54, 62, 63, 65,
 67, 82, 86, 96, 101, 114, 160
 Cartesian m.-b. p., 31
 Mind, Language, and Reality (Put-
 nam), 18
 mirror image, 157
 molecules, 9, 10, 50, 51
 behavior of m., 37, 58
 carbon-based m., 49
 DNA m., 59
 H₂O m., 50, 53, 54
 metal m., 99
 objects as collection of m., 59
 ontology of m., 122
 oxygen m., 91
 movements of m., 59, 60, 69
 monism, 1, 55. *See also* materialism,
 physicalism
 anomalous m., 22
 m.-dualism dilemma, 1–4, 9, 18, 19,
 24, 25, 33, 34, 37–39, 54, 68,
 73, 95, 101, 160, 163
 eliminative m., 21
 moods, 51
 Multiple Drafts Model, 73, 75, 78, 142
 multiple realizability, 2, 5, 18–20, 22,
 24, 62
 mutual exclusion principle, 34, 36
 The Mystery of Consciousness (Searle),
 133
 Myth of the Given, 25
 Nagel, Thomas, 12, 13, 20, 22, 41, 58,
 80, 159

Nagel, Thomas, *con't.*

“What Is It Like to Be a Bat?” 12

Nannini, Sandro, 90

natural world, 1, 5, 11, 19, 24, 26, 30,
34, 44, 47, 50, 59, 68, 90, 94,
96, 101, 138, 152, 157, 159

naturalism, 87, 115, 147, 152, 159

bald n., 17

biological n., 4, 5, 33, 38–71, 73, 94,
97–99, 101, 116, 117, 120,

128, 132, 152, 156, 160, 161

conjugability of n., 97, 139

eliminative n., 73

strong n., 24

nature, 4, 5, 13, 29, 30, 33, 41, 49, 60,
74, 91–93, 101, 140, 146, 159

NCC. *See* consciousness: neuronal
correlate of

neodualism, 14, 23, 64, 85, 95

Network, 102, 109–111, 114, 117, 118,
125, 139. *See also* Background

n. of action schemes and perceptual
categories, 109

n. of causal relations, 37

holistic N., 126, 162

n. of intentional states, 105, 110,
111, 132, 156

n. of mental states, 1, 3, 6, 126, 158

n. of perceptions and thoughts, 135

pragmatic nature of N., 156

representational n., 109

n. of semantics in brain, 126

n. of thoughts, 37

neural activity, 10

neuroanatomical connections, 137

neuroanatomy, 136

neuron(s), 7, 48, 51, 53, 55, 56, 89, 99,
133, 136, 137

n. firings, 23, 52, 53, 61, 68, 70

mirror n., 102, 127–129

neurobiological processes, 47, 50, 118.

See also biological processes

neurophysiological states, 51, 58, 115

nomological connections, 90, 94

non-rational creatures, 155

normativity, 111

noumenon, 11

objects, 6, 21, 37, 43, 47, 59, 60, 63,
70, 115, 116, 123, 124, 126,
133, 137, 147 153

ability to attend to o., 148

o. with which brain/mind interacts,
105, 149

blindsight patients, o. seen by, 133

cultural o., 13

o. of experience, 104, 157

material/physical/solid o., 47, 59,
80, 93, 123

mental patterns of an o., 147, 148

physical dimensions of o., 109, 127

physics of o., 123

solid o., 59, 93

observation, 11

ontology, 1, 9, 56

o. of consciousness, 2, 3, 14, 34, 39,
46, 47, 73, 91

first-person o. of consciousness,
93, 95, 152

non-reductive o. of conscious-
ness, 39

o. divide, Cartesian, 10

dualistic o., 85, 95

first-person o., 98

o. to which “I” refers, 154

o. of mental states/mind, 5, 6, 22,
24, 33, 38, 39, 49, 57, 58, 99,
114, 152, 155, 158, 159, 160

irreducibility of o. of mind to
third-person o., 56

non-reductive o. of mind, 68

subjective o. of mind, 56

o. of molecules, 122

neuronal processes, third-person o.
of, 93

objective/subjective o., 48, 58, 74,
83, 94, 140, 144, 152, 152,
154, 155, 156

social o., 116, 120

organism-environment transactions,
2, 146

orientation, 12, 29, 30, 110

- overdetermination, 62, 64, 68, 70
 - causal o., 3, 63, 69
- pain, 9, 13, 18, 19, 20, 22, 23, 28, 30, 31, 36, 40–46, 51, 58, 61, 65, 83, 140, 155
- particles, (micro)physical, 9, 17, 92, 93
- perception (perceiving), 6, 10, 30, 31, 40, 41, 45, 51, 64, 74, 98, 105–109, 111, 114, 115, 124–130, 133, 134, 139, 143, 145, 150, 151, 154, 155, 160
 - ambiguous p., 109
 - conscious p., 57
 - visual p., 46
- phenomenal properties, 86, 90, 159, 160
- phenotype, 54
- Philosophy and the Mirror of Nature* (Rorty), 25
- physical states, 1–3, 16, 22, 51
- physical systems, 2, 4, 17, 20, 22, 47, 62, 68, 70, 86, 94, 141, 156, 161
- physical, the, 1–4, 9, 10, 13–17, 19, 24, 28, 34–37, 47, 48, 51, 52, 58, 61, 63, 65, 67–71, 85, 86, 88–90, 93, 96–98, 160, 161. *See also* mental, the
- physicalism, 1. *See also* materialism, monism, identity theory: token
 - minimal/token p., 23, 62
 - reductive p., 5
- pineal gland, 15, 78, 140
- Plato, 1, 33
 - P.-Aristotle tradition, 119
 - P.-Christian tradition, 140
 - P. view of possible worlds, 90
- Pleistocene man, brain of, 115
- pluralism. *See* dualism: Cartesian
- Popper, Sir Karl Raimund, 1, 10, 13, 14, 33, 51, 68, 159
 - The Self and Its Brain* (with Eccles), 14
- positron emission tomography (PET), 10
- power(s), 26, 27, 117, 122
 - agent, p. treated as, 27
 - biological p., 152
 - causal (government) p. of
 - consciousness/mind, 4, 22, 23, 49, 51, 54, 56, 59, 67, 68, 70, 73, 93, 94, 132, 142, 147, 152, 153, 156, 158
 - creative p., 56
 - deontic p., 122, 123
 - God's p., 87
 - mental (of the mind) p., 2, 3, 6, 9, 27, 29, 30, 31, 50, 98
- prefrontal lobes, 36, 150. *See also* brain(s)
- preintentional skills (abilities, capacities), 105, 113, 125, 128, 132, 156, 162
- prelinguistic children/creatures/organisms, 111, 152, 155, 157
- present, remembered, 45, 135, 146
- presuppositions, 13, 33, 36, 47, 62, 63, 68, 79, 116, 120, 152, 153
 - conscious p., 95
 - theoretical p., 9, 86, 95
- process(es), 3, 19, 36, 83
 - active/passive p., 136
 - biological p., 2, 27, 28, 49, 50, 58, 151, 157, 160
 - brain p., 5, 44, 47, 52, 64, 94, 135, 160, 162
 - causal p., 2, 19, 65, 75, 161
 - cognitive p., 85, 109, 125, 142, 143, 145, 150
 - computational p., 152
 - dynamic core as p., 24, 95
 - embodied p., 4, 138, 159, 162
 - higher/lower-level p., 51, 88
 - language p., 85
 - learning p., 44
 - macro/microstructural p., 4
 - memory p., 136
 - mental/mind p., 1–4, 19–22, 24, 36, 49, 97, 101, 108, 128, 143, 146, 148, 157, 159, 160
 - multiply realizable p., 2, 19, 21, 94, 95

- neur(al)(onal) p., 50, 53, 93, 136, 137, 151
- neurobiological p., 47–50, 118
- neurophysiological p., 49
- nonmodular p., 79
- objective p., 13, 80
- ontologically irreducible p., 155
- perceptual p., 136
- physical p., 4, 35
- subjective p., 30, 138, 162
- (un)conscious(ness) p., 10, 81, 85, 94, 131, 132, 134–136, 138, 158
- promise(s), 103, 124
- propert(ies)(y), 1, 23, 61, 70, 74, 88, 116, 161
 - A-/B-p., 87, 88
 - biological p., 87
 - brain p., 11, 50, 54, 116, 136
 - conscious p., 41
 - emergent p., 49, 50, 51, 54, 55, 57, 58, 60, 62, 60, 62, 99
 - essential p., 56
 - experiential p., 41
 - functional p., 127, 138
 - (ir)reducible p., 67, 73
 - higher/lower-level p., 51, 54, 73, 88
 - macrophysical p., 14
 - material p., 134
 - mental (mind)/physical p., 3, 9, 13, 14, 23, 16, 17, 19, 22–23, 34, 35, 52, 54, 62, 63, 67, 70, 84, 90, 91, 96
 - microphysical p., 14
 - neuro(biological)(nal) p., 22, 47, 48, 50, 53, 93, 132, 134
 - p. ontologically independent of the physical, 48, 50, 96
 - P-conscious p., 41
 - phenomenal p., 86, 90, 159, 160
 - psychological 86, 140
 - qualitative p., 1
 - second-order p., 22, 23, 65, 78, 80
 - sub/supervenient p., 52, 63, 87
 - system p., 68
- property dualism. *See* dualism: property
- propositional content, 41, 42, 103, 105
- psychoanalysis, 81
- psychological mode, 102–104, 118
- psychophysical nexus, 11
- psychophysical relation, 3–5, 13, 17, 48, 50, 52, 53, 57, 65, 67, 73, 95, 99
- Putnam, Sir Hilary, 18, 20, 22, 118, 119
 - Mind, Language, and Reality*, 18
- Pythagorean theorem, 106
- qual(e)(ia), 20, 41–44, 64, 65, 79, 80, 82, 89, 95, 134, 147, 160
 - epiphenomenal q., 65
- qualitative state(s), 42, 74, 82, 83, 101, 131
- qualitativeness, 1, 34, 36, 39, 41–44, 47, 55, 74, 78, 80, 82, 85, 101, 131, 133, 138, 139, 156, 158, 159, 161
- quantum phenomena, 9
- Quine, Willard van Orman, 1, 14, 33, 75
- rationality, 1, 36, 110, 111, 116, 152, 153, 155
 - Rationality in Action* (Searle), 152
- realism, 75, 81, 115
- reality, 25, 59–61, 80
 - appearance-r. distinction, 61, 84
 - created new r., 102
 - empirical r., 45
 - independently existing r., 105
 - institutional r., 125
 - layered model of r., 48, 49
 - natural r., 67
 - non-reductive view of r., 99
 - objectiv(e)(ity) of r., 81, 97
 - physical r., 61, 97
 - selves, reality of, 142
 - social r., 124
- realizability, multiple, 2, 5, 18–20, 22, 24, 62
- realizers, 22, 23, 67
- reasoning, 41, 42, 45, 64, 66, 83, 94, 129, 140, 158
 - inferential r., 11
- The Rediscovery of the Mind* (Searle), 49, 116, 152

- reduction(ism), 16–18, 34, 43, 51, 57,
58, 67–71, 73, 82, 85, 87, 88,
95, 96, 132, 134, 157
antir., 3, 9, 12, 14, 15, 17, 23, 35,
43, 48, , 57, 58, 64, 65, 67, 87,
92, 93, 97, 98, 139, 159, 161
antir. arguments, 20, 24
biological r., 48
Putnam's antir., 22
causal r., 59, 60, 93
eliminative r., 47
functional r., 67
ontological r., 59–62, 67, 93, 94
non-r., 147, 152
r. to monism/dualism, 4
non-r., 5
strong r., 22
third-person r., 47
reentrant mechanisms, 131, 135
The Remembered Present (Edelman), 146
representation, 6, 10, 41, 43, 66, 104,
111, 113, 124
conscious r., 57
intentional r., 102, 103
mental r., 102, 110
r. Network, 109, 110, 158, 162
non-r. mental capacities, 84, 112
non-r. phenomena, 115
perception, r. theories of, 106
pictorial r., 129
res cogitans, 31, 45, 78, 80, 96, 98
res extensa, 45, 98
Revonsuo, Antti, 126, 134
Rizzolatti, Giacomo, 127–128
Rorty, Richard, 1, 14, 21, 25, 80, 83, 159
Philosophy and the Mirror of Nature, 25
Rorty-Dennett line of thought, 83
Rosch, Eleanor, 6
Rudder-Baker, Lynne, 67
rules, 21, 80, 112, 113
constitutive r., 121, 123, 124
language r., 154
regulative r., 123
Ryle, Gilbert, 75
satisfaction, 6, 43, 79, 102–112, 114,
119, 132, 156, 162
scene(s), 81, 127, 135, 137, 148, 157
culture and language, s. dominated
by, 141
mental s., 132
perceptual s., 128
unitary s., 137
schizophrenia, 81
science, 37, 44, 59, 74, 75, 87, 91,
140, 163
cognitive s., 1, 4, 18, 24, 28, 37, 64,
79, 80, 81, 85, 86
computer s., 18
epistemic s., 45
history of s., 59
linguistic game of s., 74
objectivity of s., 44, 45, 97, 159
reductionist s., 95
Searle, John R., 1, 3, 4–6, 9, 17, 20–22,
28, 33–60, 62, 67–70, 73, 74,
80–84, 90, 92–99, 101–129,
131–139, 141, 146, 147, 151–162
“Collective Intentions and Ac-
tions,” 120
The Construction of Social Reality, 120
Intentionality, 80, 162
Expression and Meaning, 102
Mind, Language, and Reality, 18
The Mystery of Consciousness, 133
Rationality in Action, 152
The Rediscovery of the Mind, 49,
116, 152
Speech Acts, 80, 102, 123
sel(f)(ves), 1–3, 6, 10, 28, 131, 132,
135, 139, 142, 146, 148, 149,
152, 153, 156, 157, 158, 160
actual s., 140
autobiographical s., 149, 151
Cartesian view of ourselves, 25
conscious s., 139
core-s., 147, 149, 150, 151
elaborate s., 150
extended s., 143, 144
formal s., 155
s.-non-s. distinction, 139

- sel(f)(ves), *con't.*
 primitive s., 132, 157, 158
 proto-s., 139, 146, 150, 151, 162
 quasi-s., 142
 rational s., 99, 154, 156, 157
 real s., 146
 Searle's s., 152, 153, 154
 transitional s., 142
 virtual s., 142
The Self and Its Brain (Popper and Eccles), 14
 self-categorization, conceptual, 135
 self-consciousness, 155, 157
 self-control, 141
 self-definition, 141
 self-recognition, 152, 157
 self-referen(c)(tiality), 106, 131, 146, 158
 causal s.-r., 6, 98, 102, 105, 106, 108, 109, 114, 118–120, 131, 132, 139, 152, 156, 157, 158, 162
 preconceptual self-r., 7
 primitive self-r., 156
 Sellars, Wilfrid, 15, 25, 110, 157
 semantic dementia, 126
 semantics, 20, 21, 35
 sensation(s), 9, 29, 30, 32, 39–42, 44, 51, 60, 74, 83, 135
 sensibility, 26, 39, 154
 sensory inputs, 18, 19, 22, 125. *See also* behavior(ists)(m)(s): outputs
 Sinigaglia, Corrado, 127, 128
Sinn(en) (sense), 105, 119
 Skarda, Christine, 114
 skills, 3, 114. *See also* abilit(ies)(y)
 biologically grounded, 109
 linguistic s., 109
 mental s., 125
 preintentional s., 132, 156
 sleep, 41, 116
 dreamless s., 40, 137
 REM s., 35
 solidity, 37, 51, 54, 58–63, 68, 69, 93, 99
 solipsism, 102, 116
 Cartesian s., 128
 methodological s., 115
 soul, 14, 27, 28, 32, 78
 s.-body interaction, 140
 Cartesian s., 2
 (im)materiality of s., 2, 32
 immortal s., 140
 nonphysical s., 140
 persistence of s. after physical death, 91
 Platonic-Christian s., 140
 speaker-agent, 118, 120
 speech acts, 46, 80, 106, 111, 142, 156
 intentional states, s. a. compared with, 104
 Searle's s. a. theory, 102, 103, 116, 120, 122
Speech Acts (Searle), 80, 102, 123
 spirit(ualism), 13, 29, 35, 81
 s. view, 87
 split brain, 76, 81, 134, 137, 142
 state(s) of affairs, 6, 21, 43, 57, 92, 93, 105–107, 116, 132
 status-functions, 124
 stereotype, 136
 stimul(i)(us), 35, 109, 117, 137, 142.
 See also behavior: output
 environmental s., 109
 input s., 56
 stimulus, poverty of, 117
 subject(s), 1, 2, 10, 15, 20, 28, 30, 32, 35, 40, 45, 74, 78–81, 83, 101, 118, 127, 131, 134, 143, 144, 145, 152, 155, 157, 158, 162.
 See also "I"
 conscious s., 134
 s. of experience, 77
 impaired s., 111
 mutual recognition between s., 128
 passive s., 25
 subjectivity, 1, 5, 26, 32, 34, 39, 41, 44, 45, 47, 55, 62, 82, 84, 110, 120, 131, 139, 157, 160, 162
 consciousness, s. of, 36, 97, 138, 147, 158
 inters., 128
 irreducibility of s., 78
 ontological s., 68, 95, 139
 passive s., 6, 50, 128

- subjectivity, *con't.*
 - primitive s., 155
 - qualitative(ness and) s., 74, 78, 80, 82, 85, 101, 131, 133, 139, 156, 158, 159, 161
 - self-transparent s., 81
- subject-predicate structure of sentences, 154
- substance(s), 13, 40, 53, 90, 119, 122, 159. *See also* dualism: substance
 - chemical s., 40
 - immaterial s., 1, 15, 29, 31
 - indexically determined s., 118
 - mental/mind s., 24, 32, 96
 - realizing s., 2, 19
 - thinking s., 140
 - traditional metaphysical concept of s., 140
 - unextended s., 32
- substantialism, Cartesian, 101
- substantialist metaphysics, 26
- success-verb, 105, 106
- supervenience, 51, 63–65, 67, 86
 - logical s., 52, 85, 87–89, 92, 93, 96
 - mind-body s., 62
 - natural s., 17, 87, 91, 94
- syntax, 20, 21, 108, 111
- system(s), 54, 98, 99, 134
 - autonomous s., 116, 117
 - behavior of s., 19, 23
 - biological s., 5, 48
 - brain s., 48, 53, 70, 134
 - causally closed/ effective s., 35, 70
 - complex s., 75, 138
 - conceptual s., 90
 - control s., 77
 - coupled s., 143
 - s. devoted to creative structure, 101
 - emergent properties of s., 50, 51
 - environment, interaction of s. with, 24
 - explanation/description of s., 69, 70
 - s. features, 52, 55
 - formal s., 21
 - functional analysis of s., 22
 - identical s., 62
 - lower-level elements, s. composed of, 52
 - (macro)(micro) levels of s., 4, 49, 54, 55, 70, 73, 93, 94, 161
 - material s., 2, 31, 32
 - mental states attributed to a s., 12, 56
 - mental/mind s., 97, 143,
 - s. lacking mental states, 20
 - motor s., 127, 128
 - s. of oppositions, 35
 - organic s. (brain), 50, 51, 156
 - perceptual s., 12
 - physical s., 2, 4, 17, 20, 22, 47, 62, 68, 70, 86, 94, 141, 156, 161
 - s. property, 68
 - s. of states, 46
 - value s., 135
 - value-category s., 46, 135, 139
- thalamocortical system, 131, 134, 135, 137
- theoretical laziness, 10
- thinking, 26, 27, 30–32
 - t. being, 54
 - conscious t., 42
 - eternal t. being, 31
 - t. extended matter, 32
 - immaterial t. being/ substance, 29, 31
 - t. matter hypothesis, 2, 28, 31, 32
 - qualitative character of t., lack of, 42
 - t. substance, 140
- Thompson, Evan, 6
- thought(s), 2, 9, 10, 28, 30–33, 40, 41, 46, 55, 56, 74, 115, 121, 122, 124, 125, 142, 143, 151, 155, 156
 - conscious t., 42, 44, 135
 - t. experiment, 12, 21, 56, 89, 118
 - Chalmer's t. experiment, 145
 - identical t., 119
 - inner t., 137
 - institutional reality, t. constitutive of, 125
 - linguistic forms of t., 124
 - Lockean view of t., 27, 33
 - microhabits of t., 76
 - neo-Humean t., 141

- thought(s), *con't.*
 non-reductive line of t., 33
 prelinguistic t., 124
 t. as qualia, 42
 rational t., 50, 110
 religiously inspired t., 140
 schizophrenia, alien t. in, 81
 scientific t., 45, 95
 second-order t., 156
 widespreadness of t., 146
 thought experiments, 12, 21, 56, 89,
 118, 145
 Titchener, Edward, 135
 token(s), 2, 22–24
 t. identity theories, 17, 19
 Tomida, Yasuhiko, 25
 Tononi, Giulio, 7, 10, 36, 57, 76, 99,
 116, 126, 131, 132, 135, 137,
 138, 139
 consciousness, T.'s theory of, 162
 T.'s dynamic core theory, 126, 131,
 135–137
 A Universe of Consciousness (with
 Edelman), 135
 transduction, 78, 79
 Twin Earth, 118, 119
 Tye, Michael, 16, 42
 type, 2, 59, 90, 103, 121
 t. identity, 115, 118, 119
 t.-identical brain states, 120
 t.-identical mentalistic con-
 tents/effects, 51, 119
 t.-identical neural configura-
 tions, 115
 t.-identical neurophysiological
 causes/ states, 51, 115
 t.-identical realizations, 114, 115
 t.-identical systems, 52
 t.-i. theory, 3, 4, 17, 22, 69
 t.-identical thoughts, 119
 mental/physical t., 19, 90
 state t., 103

 underlaborer, 25
 understanding, 27, 29, 30, 32, 60, 67,
 106, 117
 meaning, u. literal, 112
 u. mind, 96, 97
 prelinguistic u., 129
 pre-u., 39
 unity, 55, 131, 138, 140, 141, 144–
 146, 159
 u. of apperception, 45
 u. of consciousness, 39, 41, 45, 47,
 74, 76, 126, 131, 133, 138
 u. of experience, 81
 holistic, 45, 46, 82, 156, 158
 horizontal u., 46
 u. of the “I,” 45
 u. of natural world, 90
 u. of scientific view, 59
 u. of subject, 81
 synchronic u., 47, 136
A Universe of Consciousness (Edelman
 and Tononi), 135

 Van Gulick, Robert, 67
 Varela, Francisco, 6
 virtual machine, 75–77
 vision, 84, 104, 134
 color v., 12
 visual cortex, 133
 vitalism, 91
 volition, 26, 30, 104

 Weiskrantz, Larry, 133
 well-being, 2, 25, 26
 “What Is It Like to Be a Bat?” (Nagel), 12
 Wittgenstein, Ludwig, 74, 78, 80, 109, 154
 worldview, scientific, 14, 155

 Yolton, John, 32

 Zeki, Semir, 133
 zombie(s), 65, 89
 argument, 16, 17, 20, 22, 23, 55, 85,
 88, 90, 91, 94, 159, 161
 z. twin, 85, 161
 z. world, 90, 91, 93, 95, 161



VIBS

The **Value Inquiry Book Series** is co-sponsored by:

Adler School of Professional Psychology
American Indian Philosophy Association
American Maritain Association
American Society for Value Inquiry
Association for Process Philosophy of Education
Canadian Society for Philosophical Practice
Center for Bioethics, University of Turku
Center for Professional and Applied Ethics, University of North Carolina at Charlotte
Central European Pragmatist Forum
Centre for Applied Ethics, Hong Kong Baptist University
Centre for Cultural Research, Aarhus University
Centre for Professional Ethics, University of Central Lancashire
Centre for the Study of Philosophy and Religion, University College of Cape Breton
Centro de Estudos em Filosofia Americana, Brazil
College of Education and Allied Professions, Bowling Green State University
College of Liberal Arts, Rochester Institute of Technology
Concerned Philosophers for Peace
Conference of Philosophical Societies
Department of Moral and Social Philosophy, University of Helsinki
Gannon University
Gilson Society
Haitian Studies Association
Ikeda University
Institute of Philosophy of the High Council of Scientific Research, Spain
International Academy of Philosophy of the Principality of Liechtenstein
International Association of Bioethics
International Center for the Arts, Humanities, and Value Inquiry
International Society for Universal Dialogue
Natural Law Society
Philosophical Society of Finland
Philosophy Born of Struggle Association
Philosophy Seminar, University of Mainz
Pragmatism Archive at The Oklahoma State University
R.S. Hartman Institute for Formal and Applied Axiology
Research Institute, Lakeridge Health Corporation
Russian Philosophical Society
Society for Existential Analysis
Society for Iberian and Latin-American Thought
Society for the Philosophic Study of Genocide and the Holocaust
Unit for Research in Cognitive Neuroscience, Autonomous University of Barcelona
Yves R. Simon Institute

Titles Published

Volumes 1 - 156 see www.rodopi.nl

157. Javier Muguerza, *Ethics and Perplexity: Toward a Critique of Dialogical Reason*. Translated from the Spanish by Jody L. Doran. Edited by John R. Welch. A volume in **Philosophy in Spain**

158. Gregory F. Mellema, *The Expectations of Morality*

159. Robert Ginsberg, *The Aesthetics of Ruins*

160. Stan van Hooft, *Life, Death, and Subjectivity: Moral Sources in Bioethics*. A volume in **Values in Bioethics**

161. André Mineau, *Operation Barbarossa: Ideology and Ethics Against Human Dignity*

162. Arthur Efron, *Experiencing Tess of the D'Urbervilles: A Deweyan Account*. A volume in **Studies in Pragmatism and Values**

163. Reyes Mate, *Memory of the West: The Contemporaneity of Forgotten Jewish Thinkers*. Translated from the Spanish by Anne Day Dewey. Edited by John R. Welch. A volume in **Philosophy in Spain**

164. Nancy Nyquist Potter, Editor, *Putting Peace into Practice: Evaluating Policy on Local and Global Levels*. A volume in **Philosophy of Peace**

165. Matti Häyry, Tuija Takala, and Peter Herissone-Kelly, Editors, *Bioethics and Social Reality*. A volume in **Values in Bioethics**

166. Maureen Sie, *Justifying Blame: Why Free Will Matters and Why it Does Not*. A volume in **Studies in Applied Ethics**

167. Leszek Koczanowicz and Beth J. Singer, Editors, *Democracy and the Post-Totalitarian Experience*. A volume in **Studies in Pragmatism and Values**

168. Michael W. Riley, *Plato's Cratylus: Argument, Form, and Structure*. A volume in **Studies in the History of Western Philosophy**

169. Leon Pomeroy, *The New Science of Axiological Psychology*. Edited by Rem B. Edwards. A volume in **Hartman Institute Axiology Studies**

170. Eric Wolf Fried, *Inwardness and Morality*
171. Sami Pihlstrom, *Pragmatic Moral Realism: A Transcendental Defense*. A volume in Studies in **Pragmatism and Values**
172. Charles C. Hinkley II, *Moral Conflicts of Organ Retrieval: A Case for Constructive Pluralism*. A volume in **Values in Bioethics**
173. Gábor Forrai and George Kampis, Editors, *Intentionality: Past and Future*. A volume in **Cognitive Science**
174. Dixie Lee Harris, *Encounters in My Travels: Thoughts Along the Way*. A volume in **Lived Values: Valued Lives**
175. Lynda Burns, Editor, *Feminist Alliances*. A volume in **Philosophy and Women**
176. George Allan and Malcolm D. Evans, *A Different Three Rs for Education*. A volume in **Philosophy of Education**
177. Robert A. Delfino, Editor, *What are We to Understand Gracia to Mean?: Realist Challenges to Metaphysical Neutralism*. A volume in **Gilson Studies**
178. Constantin V. Ponomareff and Kenneth A. Bryson, *The Curve of the Sacred: An Exploration of Human Spirituality*. A volume in **Philosophy and Religion**
179. John Ryder, Gert Rüdiger Wegmarshaus, Editors, *Education for a Democratic Society: Central European Pragmatist Forum, Volume Three*. A volume in **Studies in Pragmatism and Values**
180. Florencia Luna, *Bioethics and Vulnerability: A Latin American View*. A volume in **Values in Bioethics**
181. John Kultgen and Mary Lenzi, Editors, *Problems for Democracy*. A volume in **Philosophy of Peace**
182. David Boersema and Katy Gray Brown, Editors, *Spiritual and Political Dimensions of Nonviolence and Peace*. A volume in **Philosophy of Peace**
183. Daniel P. Thero, *Understanding Moral Weakness*. A volume in **Studies in the History of Western Philosophy**

184. Scott Gelfand and John R. Shook, Editors, *Ectogenesis: Artificial Womb Technology and the Future of Human Reproduction*. A volume in **Values in Bioethics**
185. Piotr Jaroszyński, *Science in Culture*. A volume in **Gilson Studies**
186. Matti Häyry, Tuija Takala, Peter Herissone-Kelly, Editors, *Ethics in Biomedical Research: International Perspectives*. A volume in **Values in Bioethics**
187. Michael Krausz, *Interpretation and Transformation: Explorations in Art and the Self*. A volume in **Interpretation and Translation**
188. Gail M. Presbey, Editor, *Philosophical Perspectives on the "War on Terrorism"*. A volume in **Philosophy of Peace**
189. María Luisa Femenías, Amy A. Oliver, Editors, *Feminist Philosophy in Latin America and Spain*. A volume in **Philosophy in Latin America**
190. Oscar Vilarroya and Francesc Forn I Argimon, Editors, *Social Brain Matters: Stances on the Neurobiology of Social Cognition*. A volume in **Cognitive Science**
191. Eugenio Garin, *History of Italian Philosophy*. Translated from Italian and Edited by Giorgio Pinton. A volume in **Values in Italian Philosophy**
192. Michael Taylor, Helmut Schreier, and Paulo Ghiraldelli, Jr., Editors, *Pragmatism, Education, and Children: International Philosophical Perspectives*. A volume in **Pragmatism and Values**.
193. Brendan Sweetman, *The Vision of Gabriel Marcel: Epistemology, Human Person, the Transcendent*. A volume in **Philosophy and Religion**
194. Danielle Poe and Eddy Souffrant, Editors. *Parceling the Globe: Philosophical Explorations in Globalization, Global Behavior, and Peace*. A volume in **Philosophy of Peace**
195. Josef Šmajs, *Evolutionary Ontology: Reclaiming the Value of Nature by Transforming Culture*. A volume in **Central-European Value Studies**
196. Giuseppe Vicari, *Beyond Conceptual Dualism: Ontology of Consciousness, Mental Causation, and Holism in John R. Searle's Philosophy of Mind*. A volume in **Cognitive Science**