

The background of the cover is a classical-style painting. It depicts a man with a shocked or distressed expression, his eyes wide and mouth slightly open. He has a beard and is wearing a light-colored shirt. His hands are pressed against his head, suggesting intense mental anguish or a crisis. In the foreground, there is a large, glowing, translucent orb or sphere, possibly representing a concept like the self or a cognitive process. The lighting is dramatic, with strong highlights and deep shadows.

Michele Di Francesco  
Massimo Marraffa  
Alfredo Paternoster

# The Self and its Defences

*From Psychodynamics to  
Cognitive Science*



# The Self and its Defenses

Michele Di Francesco • Massimo Marraffa • Alfredo Paternoster

# The Self and its Defenses

From Psychodynamics to Cognitive Science

palgrave  
macmillan

Michele Di Francesco  
School of Advanced Studies  
IUSS Pavia, Italy

Alfredo Paternoster  
University of Bergamo  
Bergamo, Italy

Massimo Marraffa  
University of Roma Tre  
Rome, Italy

ISBN 978-1-137-57384-1      ISBN 978-1-137-57385-8 (eBook)  
DOI 10.1057/978-1-137-57385-8

Library of Congress Control Number: 2016955530

© The Editor(s) (if applicable) and The Author(s) 2016

The author(s) has/have asserted their right(s) to be identified as the author(s) of this work in accordance with the Copyright, Designs and Patents Act 1988.

This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made.

Cover image © Artepics/Alamy Stock Photo

Printed on acid-free paper

This Palgrave Macmillan imprint is published by Springer Nature  
The registered company is Macmillan Publishers Ltd.

The registered company address is: The Campus, 4 Crinan Street, London, N1 9XW, United Kingdom

# Acknowledgements

This book is the result of a long-lasting cooperation between the authors. It arises from the confluence of two main lines of research. The first one relates to the themes of personal identity and the place of subjectivity in the world order. The second one pertains to the fruitfulness of a bottom-up, ontogenetic approach to human self-awareness, one that attempts to reconstruct how the complex psychological functions underlying the adult self-conscious mind evolve from more basic ones.

We have focused and pursued these lines of research for many years, and we obviously accumulated a great intellectual debt with quite a number of friends and colleagues, who offered criticism and advice, skepticism and support, over such a long period of gestation.

Many people created a lively intellectual environment where the seeds of many of the ideas expressed in this book could eventually breed. In particular, we are grateful to Luciano Arcuri, Grazia Attili, Lynne Baker, Sergio Fabio Berardini, Claudia Bianchi, Clotilde Calabi, Riccardo Chiaradonna, Roberto Cordeschi, Mario De Caro, Roberta de Monticelli, Rosaria Egidi, Carlo Gabbani, Rossella Guerini, Diego Marconi, Stefano Meacci, Mario Miegge, Simonetta Montanari, Roberto Mordacci, Michael Pauen, Giulia Piredda, Massimo Reichlin, Andrea Sereni, Alberto Voltolini. Michele Di Francesco owes a personal debt of gratitude to Stefano Cappa and Andrea Moro for sharing their knowledge and thus making the interaction between neuroscience and philosophy possible and indeed fruitful.

We were fortunate enough to be able to discuss the ideas developed in the book in workshops, conferences and seminars in various venues, such as Bucharest, Cracow, Granada, L'Aquila, London, Madrid, Milan, Pavia, Parma, Prague and Rome. We are grateful to the organizers and the participants—too many to mention by name—for numerous instructive comments and for their worthwhile advice.

George Graham's initial perplexities about the project of this book pushed us to a better understanding of what our real subject was. A book in Italian (Marraffa & Paternoster, 2013) and a series of papers (Di Francesco & Marraffa, 2013, 2014; Di Francesco, Marraffa & Paternoster, 2014; Marraffa, 2014, 2015; Marraffa & Paternoster, 2016) served as preparatory sketches of the final fresco we struggled to paint in this work. We express our warm gratitude to the reviewers who helped us in the process of clarification of our thinking. Philip Gerrans' valuable comments on Marraffa (2011a) allowed a better understanding of the theory of mind debate. Neil Campbell's suggestions on a draft of Marraffa and Paternoster (2016) addressed important revisions of the issues discussed in Chap. 3. Rupert Dörflinger's sympathetic and critical comments on Marraffa (2014) paved the way to the treatment of Locke's link between self-consciousness and responsibility in the Epilogue.

Special reference needs to be made to the influence exerted by the psychiatrist Giovanni Jervis on this book. Jervis was a prominent figure in the Italian intellectual landscape of the second half of the twentieth century. Marraffa, who was his student and friend, has argued in a series of papers (2011b, 2012, 2013) that Jervis delivers us the premises of a philosophical anthropology that aims to integrate Ernesto de Martino's phenomenological psychology of identity and the psychodynamic theme of defense mechanisms into the naturalistic framework of biological and psychological sciences. The fourth chapter of our book has grown from these powerful insights.

To Alfredo Tomasetta we are grateful for his thoughtful engagement with Chaps. 2 and 4, and for an endless series of discussions on (the limits of) naturalism. Only Cristina Meini read and commented on the whole manuscript; to her our debt is deep, since practically every chapter is better for her questions and suggestions. A special thank you is due also to Stefano Bacin for his (critical) comments on our reading of Kant.

## References

- Di Francesco, M., & Marraffa, M. (2014). A plea for a more dialectical relationship between personal and subpersonal levels of analysis. *Frontiers in Psychology*, *5*, 1165.
- Di Francesco, M., Marraffa, M., & Paternoster, A. (2014). Real selves? Subjectivity and the subpersonal mind. *Phenomenology and Mind*, *7*, 118–133.
- Marraffa, M. (2011a). Theory of mind. In J. Fieser (Ed.), *The internet encyclopedia of philosophy*. <http://www.iep.utm.edu/theomind/>
- Marraffa, M. (2011b). Precariousness and bad faith. Jervis on the illusions of self-conscious subjectivity. *Iris*, *3*(6), 171–187.
- Marraffa, M. (2012). Remnants of psychoanalysis. Rethinking the psychodynamic approach to self-deception. *Humana.Mente*, *20*, 223–243.
- Marraffa, M. (2013). De Martino, Jervis, and the self-defensive nature of self-consciousness. *Paradigmi*, *31*(2), 109–124.
- Marraffa, M. (2014). The unconscious, self-consciousness, and responsibility. *Rivista internazionale di filosofia e psicologia*, *5*, 207–220.
- Marraffa, M. (2015). Mindreading and introspection. *Rivista internazionale di filosofia e psicologia*, *6*, 249–260.
- Marraffa, M., & Paternoster, A. (2013). *Sentirsi esistere*. Rome-Bari: Laterza.
- Marraffa, M., & Paternoster, A. (2016). Disentangling the self. A naturalistic approach to narrative self-construction. *New Ideas in Psychology*, *40*, 115–122.

# Contents

<b>1</b>	<b>Introduction: Setting the Stage</b>	1
<b>2</b>	<b>The Unconscious Mind</b>	9
2.1	The Mind and Cognitive Science	10
2.1.1	The Computational-Representational Mind	13
2.1.2	The Dissociation Between Mind and Consciousness	18
2.1.3	Levels of Explanation	21
2.2	The Freudian Unconscious	24
2.3	The Unconscious in Cognitive Science: A Critical Discussion	36
2.3.1	Searle Against the Cognitive Unconscious	37
2.3.2	Personal and Subpersonal in Dialectical Relationship	43
2.4	The Dynamic Unconscious in a Cognitive- Evolutionary Framework	46
<b>3</b>	<b>Making the Self, I: Bodily Self-Consciousness</b>	55
3.1	The Disappearance of the Self	57
3.1.1	The Exclusion Thesis	57
3.1.2	Selfless Minds?	61
3.1.3	Analytic Kantianism	67



3.2	The Bottom-Up Reconstruction of the Self	73
3.3	Consciousness and Self-Consciousness: The Case Against Pre-Reflective Self-Consciousness	76
3.4	The I as the Making of the Me	90
<b>4</b>	<b>Making the Self, II: Psychological Self-Consciousness</b>	<b>95</b>
4.1	The Nature of Introspection	98
4.1.1	Being Able to Say Why	100
4.1.2	Self/Other Parity or Inner Sense?	105
4.1.3	Self-Interpretation Plus Sensory Access	108
4.1.4	Remnants of Introspection	112
4.2	The Construction of the Virtual Inner Space of the Mind	114
4.2.1	Mindreading and Attachment	114
4.2.2	The Construction of Introspection in the Attachment Environment	118
4.3	The Emergence of a Continuous Self Through Time	129
4.3.1	Dissociation of the Jamesian Selves	134
4.3.2	The Thread of Life	140
<b>5</b>	<b>The Self as a Causal Center of Gravity</b>	<b>147</b>
5.1	A Baconian Approach to Defense Mechanisms	148
5.2	Construction and Defense of Subjective Identity	154
5.3	Scaling Up: Culture as a System of Defense Techniques	167
5.4	A Robust Theory of the Self	174
<b>6</b>	<b>Epilogue</b>	<b>179</b>
	<b>References</b>	<b>187</b>
	<b>Index</b>	<b>211</b>

# 1

## Introduction: Setting the Stage

Reference to the notion of self plays a crucial role in a multitude of areas in philosophy and in social and human sciences; arguably most important, the notion of self seems to be an indispensable and central concept of the common-sense view of the world. It is the concept of an entity that, despite being extremely elusive and difficult to explicate, is the most fundamental piece of our mental life, something that makes all the rest of it possible. Despite this centrality, there is no consensus on what the self is, or even on its very existence.

In this book, we offer a theory of the self (which is at the same time a theory of self-consciousness, as will be clarified over the course of the book), whose core ideas are that (1) the self is a *process*, a psychobiological system activity of self-representing, and (2) this process aims mainly at defending the self-conscious subject against the threat of its metaphysical inconsistency. In other words, the self is essentially a repertoire of psychological maneuvers whose outcome is a self-representation aimed at coping with the fundamental fragility of the human subject. It is a *constructive* process that starts in the very early stages of our life and runs unceasingly throughout our entire life.

Our picture of the self differs from both the idealist and the eliminative approaches widely represented in contemporary discussion. Against the idealist approach, we deny that the self is something primitive and logically prior: a mental entity describable as the owner of its own mental states. Rather, we take it to be the result of a process of construction that starts with subpersonal unconscious processes. On the other hand, we also reject the anti-realistic, eliminative argument that, from the non-primary, derivative nature of the self, infers its status as an illusory by-product of real neurobiological events, devoid of any explanatory role. Our approach is then both *derivative* and *realistic*.

\* \* \*

Our view of the self will be justified by a combination of philosophical arguments and data from cognitive sciences. The conceptual framework of our investigation can be described as *naturalistic*, *bottom-up*, and *systemic-relational*. Let us clarify each of these perspectives.

By ‘naturalistic’ we simply mean a framework that takes science seriously, at least in the sense that it is not possible for such a perspective to be in contrast with established findings provided by scientific disciplines. Even though we are not committed to taking our scientific view of the world as the only way to address the question of the self, we do consider recent findings in the realm of cognitive neuroscience and experimental psychology as a constraint upon it.

This brings us to the idea of a ‘bottom-up’ methodology. From Descartes’ *cogito* to Husserl’s transcendental ego, philosophy has adopted an inflationary approach to the self. One proceeds *top down*, starting from the philosopher’s introspective self-consciousness, to arrive at everything else. The subject is taken to be transparent to oneself, and the knowledge provided by the reflective awareness that the mind has of its own structure and contents enjoys a special kind of certainty, which is distinct from our knowledge of the physical world. Our book invites the reader to take the opposite path: we start from the idea of the fruitfulness of a bottom-up, ontogenetic approach, which attempts to reconstruct how the complex psychological functions underlying the adult self-conscious mind evolve from more basic ones. This approach does not appeal to our introspective self-knowledge, but rather to the results of investigations into the

gradual construction of human self-awareness: from the automatic and pre-reflective processing of representations of objects (object-consciousness), through the awareness and then self-awareness of the body, up to introspective self-awareness and then narrative identity.

Our conceptual framework, however, aims to avoid not only a top-down ontologically inflationary approach to the self, but also an overly reductionist approach which explains *everything* in terms of bottom-up neurocognitive mechanisms. This is where a contextualist and systemic perspective comes into play. Here the individual's psychological problems are investigated by putting them in the inter-individual and social context in which they arise and obtain a sense. This systemic naturalism is rooted in the Chicago school of functionalism, and is the foundation of attachment theory—namely, the psychodynamic tradition within which we will develop our theory of self-consciousness.

The result of this multidimensional approach is a theory of self-consciousness according to which two aspects of the self are to be distinguished: on the one hand, there is a *selfing* process (the 'I', in Jamesian terminology), which is a synthesis function that works mainly at the subpersonal level; on the other hand there is the product of this process: the representation of the self (in James' words: the 'Me'), which is partly open to conscious inspection. The Me, which is constantly updated by the selfing process, is in the first place bodily, then psychological. The highest developmental point of this process is the narrative self, which is one among the layers of personality. This view involves a criticism of the primacy of self-conscious subjectivity, which, far from being a primary givenness, is unveiled as an articulate construction consisting of several neurocognitive and psychosocial components. As existentialist phenomenology puts it, we do not possess an essence that precedes our existence; our 'being-there' is always the being-there of a living body operating in a physical and social context, with a history. And it will be argued that this being-there is characterized primarily by its *precariousness*. In the absence of any metaphysical guarantee, the constructed self (the Me) is perpetually beset by the risk of its own disintegration. Hence the already-mentioned defensive nature of the self, its being primarily a process whose teleology is focused on self-protection or self-defence.

As the reader can already realize from these introductory remarks, there are several strands in this book. In particular, it combines cognitive psychology, analytical philosophy and psychodynamics (not to mention some

excursions into a ‘continental’ philosophical anthropology). In a vague but (we hope) understandable sense, the result is more an exercise in the philosophy of psychology than in the metaphysics of mind—even if our naturalistic methodology renders the boundaries between epistemological and metaphysical worries somewhat vague and undefined. Nor do we propose a systematic comparison with the classical phenomenological approach. We simply follow our route from subpersonal unconscious processes to the personal conscious self-representation and in doing so we address metaphysical or phenomenological problems as they present themselves.

\* \* \*

Let us now give an overview of the structure of the book.

The second chapter is devoted to an analysis of the notion of unconscious, both in the cognitivist sense (the so-called ‘cognitive unconscious’) and in the Freudian sense. We explain why cognitive sciences focus on unconscious processes and structures, strongly diminishing the importance of the conscious level, and we determine what is alive and what is dead in Freud’s theory of the unconscious. Starting from this analysis, we argue that the strategy, pursued in cognitive science, of explaining behavior and mental phenomena with unconscious or *subpersonal* processes and structures is fruitful. However, since this approach runs the risk of overextending the scope of the concept of mind (this is the ‘mark of the mental’ problem), and of making the problem of unifying personal-level explanations with scientific explanations of mental phenomena (the ‘interface problem’) more difficult, we also make a case for a dialectical relationship between personal and subpersonal levels of analysis. In particular, we submit that certain psychodynamic constructs very close to the personal level (paradigmatically, the notion of attachment) are indispensable to an account of self-consciousness. The chapter thus ends with the development of the psychodynamic framework within which to conduct our research on self-consciousness. We focus on relational themes, especially on the forms of cognitive-affective relationality of the very young child. As is shown by the theories of object relations and attachment, physical contact and the construction of protective and communicative interpersonal structures constitute the infant’s primordial psychological needs, around which her mental life gradually takes form.

In the third chapter, we undertake our realist (neither idealist nor eliminative) view of the self, arguing that the first and fundamental form of self-consciousness is the consciousness of one's own body, taken as a whole. We start with a criticism of the 'exclusion thesis', the claim that there is no room for something like the self in the natural order—a thesis that in modern philosophy goes back to Hume's and Kant's criticisms of the Cartesian self. After having dismissed the Humean eliminative approaches to the self, we turn to a critical examination of two different approaches to the theme of self-consciousness. The first perspective is that of analytic Kantianism, a line of thought that stems from Peter Strawson's *The Bounds of Sense*; the second perspective is the project to provide a naturalistic version of the phenomenological claim that conscious experience entails self-consciousness, which has been pursued especially by Dan Zahavi.

The trouble with the former, whatever its intrinsic merits, is that it is unable to provide a genuinely *empirical* account of self-consciousness: the Kantian tradition is a form of a priori philosophical psychology, or, better, transcendental epistemology, which, insofar as it is empirically unconstrained, is incompatible with our naturalistic approach. Instead of a transcendental synthesis, we posit a *psychobiological* synthetic function: the already mentioned *selfing* process. Moreover, and as a consequence of its purely conceptual character, Kant's theory of self-consciousness hinges on a view of the human subject as originally unitary; we argue, in contrast, that the subject is primarily non-unitary and gains a sense of unity in the act of raising a bulwark against the threat of not being there.

Against the phenomenological project, we show that there is no pre-reflective or non-reflective self-consciousness that accompanies every conscious state from birth. This is an empirically void construction, ultimately still reminiscent of Kantian transcendentalism. The outcome of this discussion is that the most minimal form of self-consciousness is *bodily* self-consciousness, the capacity to construct an analogical and imagistic representation of one's own body as an entire object, simultaneously taking this representation as a subject, that is, as an active source of the representation of itself. In the last section of this chapter we begin to outline, building on James, our 'processual' view of the self: we distinguish between the self as the interminable objectivation process (the I) and the self as the multidimensional representation continuously updated by this process (the Me).

Chapter 4 is devoted to the development of the *psychological self*: an account is given of how the awareness of ourselves as subjects who are bearers of mental states is constructed from the awareness of one's own body. We show that our inner world evolves through an interplay—modulated by sociocultural variables—of mentalizing abilities, autobiographical memory and socio-communicative skills. The starting point is a critical discussion of introspection: following Freud's idea that our inner life is saturated with self-deception and bad faith, we show, based on the enormous amount of confabulation data from cognitive neuropsychology and social psychology, that our knowledge of our mental states is to a large extent inaccurate. Far from realizing that our actions are actually determined by unconscious mechanisms, we 'fabricate' rational post-hoc explanations of our behavior by means of an incomplete, partial and, in many cases, seriously defective folk theory of psychology. Thus, where Descartes saw a given essence (the self-transparent consciousness-substance), there is now something *constructed*, the product of an apparatus that allows us to partially describe, and above all narratively justify, fundamentally unconscious mental processes. With this result in hand, we focus on the ontogenesis of the inner, virtual 'theater' of the mind, arguing that the construction of an introspective experiential space occurs through the process of turning one's mentalistic skills—the ability to ascribe mental states to others—upon oneself under the communicative pressure of micro-social contexts. We will look firstly at *affective* mentalization, arguing that a positive attunement in proto-conversational infant-caregiver interactions plays a crucial causal role in the construction of the phenomenology of basic emotions. We will then examine how the construction of an inner experiential space advances under the thrust of caregivers' mind-minded talk. Finally, we turn to the most mature and cognitively demanding stage in psychological awareness, that is, the development of a narrative or autobiographical self. Here we highlight the importance of the sociocultural context: data from cultural psychology show that psychological self-consciousness is not an all-or-none phenomenon; the incompleteness of the capacity to conceptualize the existence of an inner experiential space has been observed in normal adults in pre-agricultural or pre-literate agricultural cultures.

In the fifth chapter, we put forward our central thesis about the nature of the self: the idea that the self is essentially a collection of defensive strategies aimed at coping with its lack of a metaphysical guarantee. Indeed, psychological self-consciousness, far from being a stable faculty, is a precarious acquisition, continuously under construction by the subject and constantly exposed to the risk of crisis. This precariousness is, therefore, the key to grasping the defensive nature of narrative identity. Defensiveness is immanent to human self-consciousness, since the latter constitutes itself precisely in the act of taking measures against its own dissolution. The chapter concludes with a clarification of the difference between our position and eliminative accounts (such as Dennett's) about the self. We show how our naturalistic approach to the narrative self also enables us to reject the antirealist argument that infers, from the non-primary, derivative nature of the self, a view of it as an epiphenomenal by-product of neurobiological events or, alternatively, of social (or socio-linguistic) practices. The antirealists—we will argue—disregard the essential psychodynamic component of identity self-construction. The need to construct and protect the most valid identity possible is rooted in the subject's primary need to subsist subjectively, and thus to exist solidly as a describable ego, as a unitary subject. Far from being the staging of an ephemeral self-deception, the incessant construction and reconstruction of an acceptable and adaptively functioning identity is the process that puts into place our intra and interpersonal balances, and is thus the ground of psychological well-being and mental health. Unlike Dennett's Joycean monologue, in our model self-narrative is not mere empty chatter: it is a causal center of gravity. In this sense, the psychodynamic component of our theory plays a crucial role in shaping our 'robust' (i.e., genuinely realist) view of the self.



# 2

## The Unconscious Mind

In the last 50 years, the sciences of the mind have been mostly concerned with unconscious functions. Indeed, the mental processes studied by cognitive science, such as perception, reasoning or language understanding, are not accessible to consciousness. Only their inputs and outputs (and perhaps some fragmentary parts) are. We are aware of the final results of the processes, but not of their internal dynamics. In this perspective, the unconscious is, in a way, far more important than the conscious, insofar as it is the unconscious which *explains* the abilities manifested in our behavior.

On the other hand, this emphasis on ‘hidden’ processes resulted in our losing what we are inclined to regard as the mental *par excellence*: the contents of our flow of consciousness, the phantasmagoric pattern of sensations and emotions which constitute our mental life—that is, losing our self-conscious subjectivity. But, if one is not talking about *conscious* mind, is one really talking about mind at all? The answer to this question depends crucially on what one takes the *mental* to be. This is the so-called issue of the *mark of the mental* (or *the cognitive*), recently brought to prominence in the debate on the extended mind (the view according

to which cognitive systems go beyond the boundary of the organism). After discussing the criticisms leveled by John Searle against the notion of unconscious mind, we propose a notion of the mental which is able, on the one hand, to account for the relevance of unconscious functions to understanding our mental abilities, and, on the other hand, to accord to consciousness an important role in the characterization of the mental domain. In this way, we avoid the risk that, in cognitive science, the concept of the self and the related concept of consciousness end up constituting a somewhat bothersome remnant.

Our clarification of the conscious/unconscious distinction will enable us to develop the psychodynamic framework in which to conduct our research on self-consciousness. After determining what is alive and what is dead in Freud's theory of the unconscious, we will examine a central characteristic of the development of post-Freudian psychoanalysis, namely, the focus on relational themes—especially on the forms of cognitive-affective relationality of the very young child. The development of the theories of object relations and attachment is part of this trend: here, as we will see, physical contact and the construction of protective and communicative interpersonal structures constitute the infant's primordial psychological needs, around which her mental life gradually takes form.

## **2.1 The Mind and Cognitive Science**

Our scientific knowledge of mental phenomena is today provided by *cognitive science*, a collection of disciplines that aim to explain how we are able to perceive, reason, understand language, make rational choices, plan and perform actions; in brief, all the capacities that are considered as distinctively mental.<sup>1</sup> We could say that cognitive science aspires to investigate human nature across the board. This ambitious goal reveals a crucial but controversial presupposition, that is, that *human* nature, and specifically the mind, is indeed a natural fact, and as such constrained

---

<sup>1</sup> It could be argued that talking about cognitive sciences, at plural, is more appropriate. Much depends on the importance one accords to the differences between the research programs in the field. We will not be concerned with this problem here, and we will freely use the singular and the plural form without being committed to a certain epistemological position.

by the biological laws of our species. Of course, not all human behavior rests on biology, but the challenge of cognitive science—given its vocation and method—consists exactly in widening, as far as it is possible, the naturalistic realm, in denying that our choices and actions can be exhaustively attributed to historical and interpretative factors, and thus trying to overcome the dichotomy, dear to the hermeneutic tradition, between *Naturwissenschaft* and *Geisteswissenschaft*.<sup>2</sup>

A distinctive feature of the development of cognitive science has been the continuous and significant growth of importance of neurosciences. While in the 1960s and 1970s findings about the brain played a negligible role in explaining mental capacities, they currently occupy a central position. That many readers have likely heard about neuroethics, neuroaesthetics, neuropolitics and even neurotheology is a telling indication of an ‘outbreak’ of inquiries into the brain that have led some researchers to suspect that old ideas may have been presented as novel just by changing the word ‘mind’ to ‘brain’, without bringing about any actual scientific progress. In fact, the beneficial circumstance that there exists an interplay, and, within some limits, integration, between psychology and neuroscience, does not change the conviction of several researchers that the respective subjects are very well distinguished from each other. After all, this is also the folk intuition, according to which the relation between mind and brain, however close, cannot simply be couched in terms of an identity. If we can easily distinguish brains from persons—in the common sense view the brain remains, in spite of all its extraordinary importance, a physical organ, on a par with the heart or the stomach, we have difficulty in finding a firm collocation for the mind, which cannot be identified with either a person or the brain, despite being closely linked to both. Of course, common sense is not necessarily our pole star; and here, as in other cases, science must to some extent distance itself from it. We shall have to see how, where and to what extent.

\* \* \*

---

<sup>2</sup> One of the shared assumptions in cognitive science is that, although human beings have capacities that animals do not have (but note that the opposite is also true), first of all language, there is no radical discontinuity between human and non-human natures. Of course, this is Darwin’s lesson.

Much of modern common sense about the mind comes from Descartes. According to Descartes's theory of mind, the mental dimension coincides with the conscious dimension. The mind is *res cogitans*, and thought, its defining attribute, is explicated in terms of awareness (*conscientia*). The Cartesian thesis of equating the mental dimension with whatever lay within the scope of one's consciousness is partly endorsed by common sense. 'Partly' because, under the influence of psychoanalysis, today's common sense view of the mind has incorporated the idea of unconscious mental states. However, as we see in Sect. 2.2, in this folk culture of the mind there still prevails the most evident limitation of the Freudian view of the unconscious: the unconscious is but a different kind of conscious mind, in the sense that it has a structure similar to the conscious and has the possibility to become conscious.

Moreover, in Descartes's model of the subject, the mental dimension is radically distinct from the body. The body is bound by mechanical laws, is located in space, and is decomposable into parts; by contrast, the mind is free and creative, with no spatial location, and is an indissoluble unity. The Cartesian conscious mind is the locus of personal identity and (as shown by the *cogito* argument) could persist even if the body and the external world were illusory. In other words, this idea of mind is the result of a secularization of the idea of soul: like the soul, the mind is viewed as an essence that precedes existence, namely, a set of spiritual prerogatives that are *primary*, and hence *essential*, in comparison with the *accidental* nature of people's bodily determinations.<sup>3</sup>

In the twentieth-century sciences of the mind, however, both Cartesian assumptions have been rejected. The distinction between mind and body is widely denied on ontological grounds because any mental process depends on the brain and is *realized* by the brain: human beings are evolved biochemical machines.<sup>4</sup> In addition, the mental dimension is no longer

---

<sup>3</sup> Edward B. Tylor (2010) was the first to formulate the hypothesis that the natural tendency to a spiritualistic objectification of the mind (and hence the idea of soul) is due to spontaneous rationalizing mechanisms. Later, the idea of the natural origin of dualistic thinking was pursued by other thinkers, most notably by Jean Piaget. More recently, the hypothesis has been suggested again in psychology (Barrett, 2004; Bloom, 2004; Boyer, 2001) and in anthropology (Astuti, 2001; Cohen & Barrett, 2008).

<sup>4</sup> Obviously in the philosophical debate dualistic positions are still present—even if very rarely they take the form of Cartesian substantial dualism. In any case the 'mainstream' ontology of cognitive science radically denies immaterial entities.

confined to the conscious one since most of the phenomena and processes that are taken as mental by cognitive sciences are not conscious. In order to fully understand this overturning of the Cartesian approach, with special regard to the dissociation between mind and consciousness, we need to dwell on two epistemological assumptions that played a crucial role in the development of the current sciences of the mind: (1) the idea (ascribable to Alan Turing) that mental processes have a *computational* nature; (2) the idea (ascribable to Noam Chomsky) that intelligent behavior is mediated by *mental representations*. These two assumptions, together with skepticism about introspection, imply the claim at the core of our discussion in this chapter: *the dissociation between mind and consciousness*.

### 2.1.1 The Computational-Representational Mind

Historically, one of the most important arguments against the possibility of conciliating the ordinary and the scientific images of the mind is the idea that no merely mechanical system could ever show genuine intelligence. No wholly mechanical system—the idea runs—can show flexible, open-ended, creative intelligence: do something truly new, respond intelligently to the unexpected. But humans do have such capacities! So, for much of the history of modern philosophy, the prevailing wisdom was that the human mind is not merely a complex physical mechanism of some kind (though animal minds might be).

Over the last 100 years or so, however, this view has been increasingly seen as untenable. In particular, Alan Turing's seminal work on the mechanization of intelligence seems to refute the above-mentioned argument for irreconcilability. For this work—and the computer science and AI that have flowed from it—seems to show that even activities that we would consider as intelligent and creative, such as reasoning or language understanding, are within range of a machine, of a purely mechanical device. The leading idea is that a particular intelligent task can be accomplished mechanically if it is decomposed into a sequence of elementary steps, each of which is well-defined, completely specified (without ambiguity) and sufficiently basic to be readily carried out by any 'executor' whatsoever.

Think of a cooking recipe for dummies, where nothing is taken for granted, including even such an obvious instruction as ‘turn on the stove’ or ‘put the pan on the fire’.

These kinds of procedures are termed *computations* (or *algorithms*). More precisely, the concept of computation is the logical-mathematical formalization of the intuitive concept of procedure. But for our purposes, we need not be too rigorous. Just think of a computation as a computer program—a connection anyone even slightly familiar with computer science will already have made. Computers are able to perform in an intelligent way because they are programmed machines; in other terms, any intelligent activity can be accomplished through a proper sequence of basic operations: the ‘right’ program. So the idea is that mental processes can be characterized as computations, as computer programs.<sup>5</sup>

Computations, of course, work on data. So, the hypothesis that mental processes are computations requires that different kinds of information (visual, auditory, linguistic, etc.) be encoded or *represented* (see below) ‘in our head’ in some format suitable to their being processed. Like a computer program, a mental process processes input information and outputs other information. Of course, this is not to claim that the brain literally works like a computer—which is simply false—but that the processes realizing our cognitive capacities can, *at a certain level of abstraction*, fruitfully be modeled on computational processes. We will examine this point further in Sect. 2.1.2.

Therefore, saying that a mental process is a kind of computation is the same as saying that it is an information-processing process. The data on which computer programs operate need not be numerical: they can concern any domain of knowledge provided that the relevant information is *encoded*, that is, expressed in a description that can be understood by the executing system, for example in some programming language. Likewise, in the case of the mind, information concerning our bodies and the world around us must somehow be encoded in order to be processed

---

<sup>5</sup>The conception of the mind as a computational device was already put forward by Hobbes and Leibniz. But only with Turing did this intuition become a sound, grounded hypothesis able to foster a serious research program.