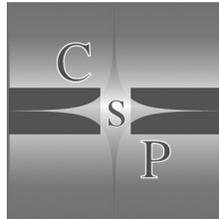


Explaining the Mental

Explaining the Mental
Naturalist and Non-Naturalist Approaches
to Mental Acts and Processes

Edited by

Carlo Penco, Michael Beaney,
Massimiliano Vignolo



CAMBRIDGE SCHOLARS PUBLISHING

Explaining the Mental: Naturalist and Non-Naturalist Approaches to Mental Acts and Processes,
edited by Carlo Penco, Michael Beaney and Massimiliano Vignolo

This book first published 2007 by

Cambridge Scholars Publishing

15 Angerton Gardens, Newcastle, NE5 2JA, UK

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

Copyright © 2007 by Carlo Penco, Michael Beaney, Massimiliano Vignolo, and contributors

All rights for this book reserved. No part of this book may be reproduced, stored in a retrieval system,
or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or
otherwise, without the prior permission of the copyright owner.

ISBN 1-84718-297-6; ISBN 13: 9781847182975

TABLE OF CONTENTS

Introduction	vi
--------------------	----

PART I: LANGUAGE AND THOUGHT

Andrew Woodfield, <i>Public Words Considered as Vehicles of Thinking</i>	2
Andrea Bianchi, <i>Speaking and Thinking (A More Kaplanian Way to a Unified Account of Language and Thought)</i>	13
Stefano Predelli, <i>The Strange Case of the Missing Constituent</i>	33

PART II: KNOWLEDGE, BELIEF AND ACTION

Pascal Engel, <i>Taking Seriously Knowledge as a Mental State</i>	50
Carlo Gabbani, <i>Epistemology and the Eliminative Stance</i>	72
Jennifer Hornsby, <i>Knowledge, Belief and Reasons for Acting</i>	88
Wolfgang Kühne, <i>Some Varieties of Deception</i>	106

PART III: INTENTIONALITY

Sandro Nannini, <i>Intentionality Naturalised</i>	124
Elisabetta Sacchi, <i>Thought and Thinking: the Ontological Ground of Intentionality</i>	135
Elisabeth Pacherie, <i>Is Collective Intentionality Really Primitive?</i>	153

PART IV: NATURALISM

Marcello Frixione, <i>Do Concepts exist? A Naturalistic Point of View</i>	178
Tim Crane, <i>Cosmic Hermeneutics vs. Emergence: the Challenge of the Explanatory Gap</i>	192
Achim Stephan and Robert C. Richardson, <i>What Physicalism Should Provide Us With</i>	207
Mario De Caro, <i>The Claims of Naturalism</i>	222

INTRODUCTION

CARLO PENCO, MICHAEL BEANEY,
MASSIMILIANO VIGNOLO

Most of the papers collected in this volume were given, in an earlier form, in a conference on mental processes held at the University of Genoa in October 2005¹. They have been supplemented by further papers from people who were involved in the conference. The aim of the volume is to present different philosophical perspectives on the mental – exhibiting, in particular, the contrast between naturalistic and non-naturalistic approaches.

There is a long tradition in philosophy of clarifying concepts such as those of thinking, knowing and believing. The task of clarifying these concepts has become ever more important with the major developments that have taken place over the last century in the human and cognitive sciences – most notably, psychology, sociology, linguistics, neurophysiology, AI, and cognitive science itself. In all these sciences, there is a need to delineate the domain of the mental and to elucidate the key concepts and underlying assumptions. This need is widely recognized, but approaches and answers vary significantly. Some stress the representational features involved in most of our mental processes, others the inferential dimension; some stress the necessity of using empirical data, others the need to refine ideas before pursuing and drawing on empirical research.

The papers collected in this volume offer a variety of perspectives on how to define, explain and understand mental acts and processes, furthering the debate among philosophers and between philosophers and those working in the human and cognitive sciences. The volume is divided into four Parts, the first on language and thought, the second on knowledge, belief and action, the third on intentionality, and the fourth on

¹ MIUR Research Project 2003110131. This research group in Analytic Philosophy brought together scholars from the Universities of Bologna, Florence, Genoa, Padua, Rome, Siena, Trieste, and Vercelli.

naturalism. In what follows we give short introductions to the topics of each Part and to the individual papers.

I. Language and Thought

The question of the relationship between language and thought is one of the perennial problems of philosophy. As two of the contributors to Part I point out, the idea of thinking as inner (voiceless) speech can be found in Plato's dialogue, the *Sophist* (263e), and the idea has been influential throughout the history of philosophy. Certainly, the idea is attractive, since it makes thinking less mysterious than it might otherwise be. Since language is more open to view, it might seem, it is more easily explained or understood than thought. If the latter can be reduced to the former, then an explanation of thought becomes possible.

One such view on the priority of language over thought emerged from Gottlob Frege's work on logic around the turn of the twentieth century. This work has often been taken to suggest, most notably, by Michael Dummett (1981; 1991, ch. 15), that language has *explanatory* priority over thought: we cannot explain thought except through an analysis of language. Even if this view has since been questioned, both as an interpretation of Frege and as correct in itself, it has been highly influential, prompting extensive research on the structure of sentences as a means of understanding the structure of thoughts. Associated with this view has also been the distinction between *thought* as the objective content of a mental process, and *thinking* as the mental process itself. Thoughts or "thinkables", as John McDowell (1994) calls them, thus become extruded from the mind. The analysis of thought as objective content of the mind becomes partly independent of the analysis of thinking as a mental process. Thoughts or thinkables are what can be expressed by language and grasped through our mental process of thinking.

This stance, motivated by the so-called "linguistic turn" that took place in philosophy in the first half of the twentieth century, has faced objections on two main fronts. First, the idea of a representational theory of the mind, where thought is considered as a direct product of a mental process, has suggested that thought need not require linguistic expression. Second, appreciation of the relevance of extra-linguistic context in determining the content of thought has raised questions about the precise relationship between language and thought. The papers in Part I of this volume address some of the issues relating to these two objections.

(i) The representational theory of mind

One of the main theories in contemporary cognitive science is the Representational Theory of Mind (RTM), which has particular implications for the question of what form the *vehicles* of language and thought take. According to RTM, mental states are mediated by internal representations, although different accounts have been given of what these are. According to Jerry Fodor (1975; 1987), one of the main advocates of RTM, these representations are language-like, that is, are given in an internal “language of thought”, which is realized in the brain. Incorporating functionalism, his claim is that what counts is the structure of these representations, which can be implemented in different physical media, not only in the brain but also in whatever hardware is capable of running the appropriate software. According to Fodor, these representations are written in an internal “mental language” or “mentalese” which is more basic than public language and not reducible to it. Mental representations, therefore, constitute the proper vehicle of mental contents (thoughts), not words in public language. We thus have a *duplication* of vehicles of thoughts: internal representations are vehicles of mental contents and physical states are vehicles of mental representations. In “Public Words Considered as Vehicles of Thinking”, ANDREW WOODFIELD criticizes this apparent duplication of vehicles of thought, which can be seen as reflected in the ambiguity of the term “carry”: mental representations carry content in a different way to the way that neural states carry mental representations. Neural states physically *realize* mental states, while mental states *possess* intentional content.

Discussion of the language-of-thought hypothesis, which lies at the heart of RTM, has tended to obscure the basic intuition that words in public language are vehicles of thoughts. Woodfield remarks that the arguments in favour of this hypothesis support only the conclusion that not all thinking is mediated by words. It may be true that the most primitive kinds of thinking are not mediated by words, but this does not imply that there are *no* thoughts that presuppose the organization of a public language. RTM is a powerful and influential paradigm, however, and the intuitive idea that public words are vehicles of our thoughts has recently been translated into the claim that *images* of public words rather than words themselves may serve as the vehicles of conscious thought. This claim is defended, for example, by Peter Carruthers (1996) in explaining what he calls “the introspective datum” – that we frequently experience ourselves mentally rehearsing sentences as we think. Woodfield raises a number of objections to this hypothesis. He questions, for example, how the appeal to an image is supposed to help here. For what makes an image

of a sentence an image of the relevant thought? Carruthers seems to just assume that it carries the right semantic and syntactic properties – that it is, as Woodfield puts it, an “iconic” representation. In any case, he suggests, there is an alternative story that explains the introspective datum at least as well, namely, that people acquire the *habit* of imagining sentences, through such common situations as preparing to speak in public.

Woodfield’s conclusions are intentionally modest: there may be many ways to interpret the claim that words are the vehicles of thoughts, but none of the arguments considered in his paper seem sufficient to establish it. At any rate, attempts to give substance to the claim inside RTM seem problematic, although there may be other ways to do so, as we will see shortly in relation to the second paper of Part I.

(ii) The relevance of context

Context-dependence constitutes a strong challenge to the idea that thought is determined solely by language. For wherever there is context-dependence, what thought is expressed by a sentence on a given occasion cannot simply be read off from the sentence itself, as Frege himself recognized (1918). There may be indexicals or demonstratives involved, such as “I”, “it”, “this”, “here” and “then”, whose reference can only be determined by understanding the context, and there may be more subtle forms of context-dependence, such as that exhibited in the use of names and tense. Sometimes the indexicality or context-dependence can be cashed out to some extent by offering another form of words, but as the work of John Perry (1993) and others has shown, there are cases of apparent “essential indexicality”.

The issue of context-dependence has thus been at the centre of discussion of the relation between language and thought. What semantic account can be given of linguistic expressions with a high degree of context sensitivity, such as indexicals and demonstratives? Some authors (e.g. Travis 2001; Recanati 2004) propose a radical contextualism, according to which, to understand the thoughts expressed, we need first to draw on our general cognitive abilities, rather than merely linguistic ones, to grasp the relevant contextual features: semantic interpretation presupposes pragmatic constraints. Other authors have followed a more traditional path, according to which semantics has a higher level of autonomy. This attitude has been encouraged by the formal semantic treatment of indexicals and demonstratives developed by David Kaplan (1979, 1989). Central to this project is the concept of “character” or linguistic meaning: the character of an indexical or demonstrative is a

function from the context of utterance to the content, that is, to what the person is speaking of. From the point of view of the semantic theorist, language still plays a central role in determining the thoughts expressed, despite the dependence on context. All sentences can be given a minimal semantic interpretation (in terms of truth-conditions) guided by linguistic rules (cf. Cappelen and Lepore 2005). The other two papers in Part I of this volume are concerned with issues relating to context-dependence and Kaplan's semantics.

In "Speaking and Thinking: towards a Unified Account", ANDREA BIANCHI takes up the question, recently discussed by Joseph Almog, of whether a unified account of language and thought is possible. He agrees with Almog that we should pursue this project, but rejects criticisms that Almog makes of Kaplan's approach to language and thought. Bianchi argues, on the contrary, that Kaplan's views make such a project not only possible but also worth undertaking. Bianchi's paper is divided into two sections: in the first, he considers the question of the existence of vehicles of thinking; and in the second, he contends that the vehicles of thinking are the very same as the vehicles of speaking, namely, words.

As far as the first question is concerned, Bianchi offers his own answer to the question of what it means to treat words as vehicles of thinking. Words, he claims, are both constitutive of the relevant act of speaking and contribute to determining what the act of speaking is about and (where appropriate) its truth-conditions. In these respects, he argues, acts of thinking are similar, and from the point of view of the unification project, should certainly be treated as requiring vehicles in the same sense in which speaking does. This raises the question of what these vehicles are in the case of thinking, and Bianchi addresses this question in the second section of his paper. It is here that he draws on Kaplan's recent work, and in particular, on his "consumerist" semantics, which, unlike "subjectivist" semantics, treats the words we use as semantically "prepackaged". He endorses Kaplan's claim that "we succeed in thinking about things in the world not only through the mental residue of that which we ourselves experience, but also vicariously, through the symbolic resources that come to us through our language" (p. 27). As Bianchi notes, if this is right, then words are vehicles of at least some of our thoughts. He is inclined to go further than Kaplan, though, in claiming that *all* thinking requires words as vehicles (except in the case of animals and pre-linguistic children).

As suggested above, one way in which language and thought are often seen as coming apart is in cases where contextual features of an utterance contribute to the determination of its truth-conditions. Whether an utterance of "It is raining" on a particular occasion is true or not depends

on *where* the sentence is uttered (or on the location to which the speaker is referring in using the sentence). What thought is expressed is thus also dependent on the context. The language used in such cases *underdetermines* the thought expressed. In the terminology used by STEFANO PREDELLI, the truth-conditions for such utterances involve unarticulated truth-conditional constituents. This is what Predelli labels as P1, the basic premise of what he calls “underarticulationism”:

P1 The truth-conditions for [an utterance] *u* involve unarticulated truth-conditional constituents.

An utterance of “It is raining”, for example, is true or false only once a place is fixed, but there is nothing in the structure of the sentence that makes reference to places.

P1 is often taken to threaten traditional semantic theory: if we have unarticulated constituents, then how can we develop a proper semantic theory without a more general theory of the interrelation between speakers and the world? In “The Strange Case of the Missing Constituent”, however, Predelli argues that the threat posed by underarticulationism can be resisted, for its derivation rests on a further assumption, which is false. He formulates this assumption as follows:

P2 If the truth-conditions for [an utterance] *u* involve unarticulated truth-conditional constituents, then empirically adequate interpretive systems must allow for unarticulated intensional constituents within the analysis of *u*.

P2 can be rejected, Predelli argues, once we avoid a simple-minded conception of empirical adequacy, and appreciate, in particular, the relationship between our intuitions about particular (actual or imagined) cases and the relevant intensions, construed as functions from what he calls “points of evaluation” to truth-values, as delivered by an interpretive system. Although the truth-conditions for an utterance might be underdetermined by its semantic structure, its intension is not. He accepts, however, that such a view comes at a certain cost – the cost of admitting that meaning does not (fully) determine truth-conditions. But he denies that this shows that semantic theory is rendered useless; rather, it just shows that the role and scope of semantic theory needs to be properly understood.

(iii) Language, thought and normativity

The conclusions that can be drawn from these discussions of RTM and the relevance of context do not imply that language is *not* required for the expression of thought, but they do suggest that there is a certain gap between language and thought. In fact, it is not just context-dependence that causes problems in this respect. We frequently use words in a loose or less than ideal way: in our actual sentences, there may be ambiguity, vagueness, slips of the tongue, malapropisms, misdirected references, ill-formed syntax, misleading expressions, poor formulations, and so on. When corrected, or on subsequent reflection, we will say things like “Yes, that is what I meant” or “My thought was this ...” Further sentences may offer partial clarifications, or remove particular potential confusions, without going so far as to offer a perfect reformulation. Indeed, when challenged or questioned, we may be hard put to say exactly what our thought was. All this suggests that there is indeed a gap between the language we use and the thought we express – at least, in one important sense of the word “thought”.

If this is right, then specification of the thought expressed cannot simply be a matter of identifying some neural configuration or physical process; it has an essentially *normative* dimension. What thought we attribute to someone, or what thought we take ourselves to be expressing, is dependent on interpreting the words that are used in the light of the context; and interpretation is subject to rational constraints. What thought is expressed is not simply a function of the language used (and certainly outstrips what a semantic theory can deliver), but is also governed by wider matters such as the beliefs of the speaker and the underlying conceptual practices.

We might also wish to distinguish between expressing and grasping a thought: on many occasions, it seems, we can express thoughts we do not fully or properly grasp. So even if there were some way of specifying what went on “in the mind” of someone when they uttered a sentence at a particular time, that may be insufficient to determine the thought expressed. Considerations like these suggest that it may be misguided to expect any straightforward naturalistic reduction or explanation of thought. At any rate, they certainly present a challenge to naturalistic programmes.

Even if our theories of thought draw upon the results and empirical findings of the natural sciences (and psychology and neuroscience, in particular), the contribution of philosophers is thus clearly essential. This holds not just in the case of understanding the relation between language and thought, but also in investigating the variety of mental processes and

their complex relations to action. This is demonstrated further in the papers that make up the other three Parts of this volume.

(M.B.)

II. Knowledge, Belief and Action

The study of human cognition is central to psychology and artificial intelligence. Psychologists speak more often of “cognition” than of “knowledge”, as if to stress the difference between the *mental processes* of understanding and the *concept* of knowledge. Willard Van Orman Quine (1969) attacked Rudolf Carnap’s attempt to give a philosophical foundation of knowledge based on a theory of sense data and promoted a naturalization of epistemology; after Quine there have been many attempts to develop a naturalistic account of knowledge, grounded on neurophysiological and psychological studies. That notwithstanding, we see a flourishing of philosophical theories of knowledge, both formal (like epistemic logic or theories of belief update) and conceptual (like causal or reliabilist theories of knowledge). The papers collected in Part II deal with conceptual problems concerning both the definition of knowledge and the relation between knowledge and action.

Philosophical clarification of the concept of knowledge has typically focused on three conceptual distinctions:

- (i) knowledge and belief;
- (ii) knowledge by acquaintance and knowledge by description;
- (iii) knowing how and knowing that.

We will say something about each of these in turn.

(i) Knowledge and belief: the standard view

The distinction between knowledge and belief is one of the most basic and traditional ones in the theory of knowledge. Plato, in his dialogues, especially *Meno* and *Theaetetus*, is considered the first to have drawn a distinction of this kind: *episteme* (knowledge or true belief with an account) is different from *doxa* (belief or true opinion), because it permits a firm grasp of the objects of knowledge; here “firm grasp” or “account” has been interpreted as a justification or proof. Recently, a stereotypical version of Plato’s ideas has been spread through the philosophical community as the JTB conception of knowledge: knowledge is Justified True Belief. In order to know something, it is not enough that we believe it; we also need to be justified in believing it and what we believe must be

true in order for our belief to count as knowledge. Summarizing, it is claimed that:

s knows that p
 iff
 s believes that p
 s is justified in believing that p
 p is true

We may call it the “standard view” of knowledge. The standard view has been challenged in different ways. The three main challenges concern (1) the irrelevance of the notion of “justification” for defining (some cases of) knowledge; (2) the insufficiency of the three conditions for defining knowledge; (3) the irrelevance of the notion of “belief” for defining knowledge.

The first kind of challenge comes from Ludwig Wittgenstein. Although the notion of justification is a central aspect of Wittgenstein’s philosophy, Wittgenstein suggests that we do not always need a justification in order to be said to “know” something. There is a kind of brute knowledge of what we do not doubt: in fact doubt is possible only against the background of certainty. In his remarks *On Certainty* written at the end of his life, Wittgenstein anticipated a stance familiar in the history and philosophy of science in the second half of the twentieth century, with the influence of Kuhn’s work on scientific revolutions: scientific knowledge works always inside certain paradigms which are simply assumed by scientists as the background against which scientific discoveries and problems take place. Wittgenstein’s account is characterized by a conceptual analysis of our use of the word “to know”, and treats knowledge not as an internal mental state or process, but as an objective and manifest capacity. Certainly, there are processes accompanying our knowledge, but this is not seen as relevant to the analysis of the *grammar* of knowledge. This anti-mentalistic point of view is not incompatible with scientific studies of the mental processes underlying knowledge, but it does open up a certain divide between philosophical analysis and scientific investigation. This divide has come under increasing attack in recent years, as philosophers have become more fully engaged in detailed discussion of mental processes, as stressed in this section by PASCAL ENGEL. A typical example of what may be called the “mentalistic turn” is given by the work of John Searle (1983), who treats belief, desire and intention as mental acts inside a general theory of intentionality. The Wittgensteinian notion of background is maintained; the background is a set of capacities that enable

the organization of the network of our mental states like beliefs and intentions. Inside this new mentalistic setting new problems arise, like the problem of the existence of “collective” intentions. We will consider this topic, discussed also by ELISABETH PACHERIE, in the next section.

The second challenge comes from the debate developed around a short paper by Edmund Gettier (1963). Gettier gave some examples showing cases in which we have justified true belief, but where we find it difficult to attribute knowledge to the subject. Let us consider a simple Gettier-style example: I have left three bottles of beer in the fridge. Some friends have drunk those beers, but Ann has bought three other bottles and put them in the fridge. When I come back home I claim to know that there are three bottles of beer in the fridge. I believe this, it is true that there are three bottles in the fridge, and I am justified in believing this because I put three bottles there. But we feel uncomfortable about attributing me knowledge in this case, because of the lack of connection between my justification and the actual causal history. Some authors have therefore claimed that knowledge is not only Justified True Belief, but that some further qualification is needed which appeals to something like causal relations or the reliability of the process of justification. After Gettier’s challenge there have been many attempts to recover the standard view, trying to supplement it with some specific requirements: Fred Dretske (1980) requires causal connections, Robert Nozick (1981) requires a subjunctive capacity to keep track of the truth, Alvin Goldman (1986) requires reliable procedures, and Hilary Kornblith (2002) requires a naturalistic account of knowledge, for example. Nevertheless, all these attempts to revise the traditional definition of knowledge have failed to achieve a satisfactory analysis, and Gettier problems make it uncertain whether the standard analysis of knowledge can be revised at all.

The third challenge to the standard definition is given by Timothy Williamson (2000) and discussed by PASCAL ENGEL in “Taking Seriously Knowledge as a Mental State”. The main claim is that it is impossible to isolate a purely “internal” element of knowledge, as if knowledge were a kind of belief with some “super” conditions. Knowledge, on the contrary, is a *factive* state, like seeing. If x sees that p , then p ; and if x knows that p , then p . The advantage of this analysis is twofold. On the one hand, we avoid the criticism made by Wittgenstein of the idea of the privacy of mental processes and states: we are allowed to speak of knowledge as a mental state not as an internal condition of an individual, but as a mental state which - being factive - is essentially connected with the external features of reality. On the other hand, we are free to abandon the analysis of knowledge in terms of justified true belief. Actually the order of

explanation is inverted: belief is explained in terms of knowledge, as the state that aims at knowledge. Engel's paper is mainly devoted to discussing this aspect of the relation between knowledge and belief; he relies upon some empirical results worked out in psychology which claim that we may attribute the idea of knowledge to children at an early stage (contrary to a shared empiricist tradition stemming from Quine). Given that children cannot be said "to believe" at the first stages of development, we are driven to accept the independence of knowledge from belief.

(ii) knowledge by acquaintance and knowledge by description

Many languages have two main verbs for "to know", like French (connetre, savoir), Italian (conoscere, sapere), Spanish (conocer, saber), Dutch (kennen, weten), German (kennen, wissen). Even if, as Saul Kripke (1977) suggested, this is not necessarily a sign of two different concepts, certainly there is something in this widespread linguistic convention. English has no such a distinction, and this lack of terminology may have encouraged discussion of the different uses of the verb "to know". We may mean knowing as being acquainted with somebody or knowing as mastering a propositional description, for example. Bertrand Russell (1910-11) dubbed the two concepts "knowledge by acquaintance" and "knowledge by description"; he thought that they are two independent ways of knowing which are not reducible to one another. Other authors have thought - partly following Kripke's claim - that we cannot have two different concepts of knowledge, but only a unifying one.

However, something analogous to Russell's distinction, in different formulations, seems to find support in empirical findings in neurophysiology, where it seems reasonable to distinguish two different functions in the brain-mind. Brain-damaged people who master direct connections between words and objects cannot give any explicit definition, while people able to give linguistic definitions of meaning may be totally unable to recognize the corresponding object. These results match our everyday experience: people may be able to recognize different kinds of objects but be ignorant of their proper definition, while a scientist who has studied the scientific definitions of some natural kind, may be unable to recognize its instances in a living situation. We are therefore prompted to distinguish two aspects of the semantic competence which guides our knowledge: referential competence which is the ability to recognize objects of a given kind when we use the corresponding word (or the ability to use indexicals), and inferential competence which is the ability to place our concept into a net of inferential relations (cf. Brandom, 1994; Marconi, 1997).

We may then provisionally accept a revised form of the Russellian distinction. One of the most famous discussions of the irreducibility of the two kinds of knowledge came from a thought experiment, the case of Mary, the scientist who has supposedly a *complete* propositional knowledge of what is red - all physical definitions in terms of wave length - but who has never seen the colour red because she has spent her entire life in an isolated room. When she comes out and sees for the first time the colour red, what does she learn? Which kind of improvement of her knowledge is realized? CARLO GABBANI, in “Epistemology and the Eliminative Stance”, discusses this kind of problem, connecting it with reductionism and defending an anti-naturalistic stance. He brings into the discussion the distinction between scientific knowledge and common knowledge, following the contrast between the manifest and the scientific images of the world discussed by Wilfrid Sellars. From his point of view what Mary lacks is not only referential competence or indexical knowledge, but a proper kind of phenomenal knowledge (the terminology comes from David Chalmers). *Phenomenal knowledge* is not a mere “private” knowledge, given in a mentalistic language. We need to postulate a fundamental similarity among the common experiences of people, as a condition of possibility of any mentalistic language and knowledge. Here “common experience” should be read as a capacity that can be described at a common sense level, and not necessarily at a neurological level. Indeed the same phenomenal knowledge might match different neurological events or states, therefore making empty the claim that a direct introspection of brain states might make phenomenal knowledge useless.

(ii) knowing how and knowing that

In our common terminology knowledge is also mastering some kind of behaviour or action. To know is not only propositional knowledge nor ability to matching names and objects, but it is also a practical skill. We all know *that* a snowboard is a board on which people move over snow, but we may not know *how* to use it. This is the first form of intuitive connection between knowledge and action, in a form that has become common in philosophy after Gilbert Ryle first distinguished between *knowing that* and *knowing how*. This theory has been sharply criticized by Williamson, who claims that *knowing how* is reducible to propositional knowledge. A partly different answer concerning the viability of the distinction is given in the paper by PASCAL ENGEL included in this Part. Engel, relying on recent results in evolutionary psychology, uses the idea of “core knowledge” - defined by psychologists like Susan Carey and

Elisabeth Spelke. Core knowledge is supposed to be a common mechanism in the brain underlying different kinds of knowledge. Engel claims that “core knowledge” is a layer (or a kind) of knowledge which is intermediary between the “how” and “that”. He therefore argues that we should abandon the division of the two kinds of knowledge, looking for a new setting for discussing the concept.

Apparently, the discussion of the relation between knowledge and action cannot be constricted in the small cage of the distinction between knowing how and knowing that. One of the main aspects of knowledge is its role in *explaining* actions. We may explain actions describing the physical mechanism of our body, but we normally explain action by giving *reasons*. How can we explain the causal efficacy of knowledge (reasons) concerning action? Going against a traditional distinction between reasons and causes, which goes back at least to Frege and Wittgenstein, Donald Davidson (1963; 1970) tried to collapse reasons and causes, saying that reasons are causes of actions. The topic has been widely discussed and a central point of the discussion is the role of the belief-desire-intention (BDI) model of folk psychology, that is, the standard conception of the relation between belief, desire and intention. This view is often considered as giving the backbones of a philosophy of action: actions are caused by desires and by beliefs that the desires will be satisfied by a certain course of action. JENNIFER HORNSBY, in “Knowledge, Belief and Reasons for Acting”, criticizes this view, arguing not only that knowledge can provide reasons for acting but also that failure to recognize this results in a mistaken view of acting for reasons. She begins by distinguishing two kinds of reasons for acting: *objective* reasons, invoked in explanations of the form “ $x \Phi$ -d because p ”, and *subjective* reasons, invoked in explanations of the form “ $x \Phi$ -d because she believed that p ”. She criticizes attempts to understand the former in terms of the latter, but does not do so by rejecting altogether the belief/desire account of action, which often underlies such attempts. Rather, she argues that knowledge itself should be seen as capable of providing reasons for acting, a point that might be expressed by taking “ $x \Phi$ -d because p ” as equivalent to “ $x \Phi$ -d because x knew that p ”. When someone acts because they believe that p , she writes, “the thought that p plays the role that the fact that p plays for someone who acts because they know that p ” (p. 102). She goes on to elaborate her view of the traditional belief/desire account, arguing not that it contains error but that it fails to give the whole truth. She rejects the distinction between motivating and normative reasons that is often associated with such an account, this distinction being part of a broader naturalistic vision which she also urges should be abandoned.

The complexity of the relationship between knowledge, belief and action is also exhibited in the final paper of Part II, entitled “Some Varieties of Deception”. WOLFGANG KÜNNE presents here a detailed analysis of the semantic field of deception, drawing on the accounts offered by Bernard Bolzano and Gottlob Frege. He starts with Bolzano’s definition of deception, which he formulates as follows:

- (D) x *deceives* y in regard to P if and only if
- (1) x brings it about, by acting as he does, that y acquires, or retains, the belief in P ; &
 - (2) P is false.

He notes that this allows for unintentional deception, and modifies (D) to define intentional deception. He comments on the related concept of misleading someone, and defines both concept of cheating, as a special kind of deceiving, and the trickier concept of intentional self-deception.

In the second and third sections of his paper, Künne turns to the concept of trying to deceive. He first characterizes the general notion, before defining, in turn, simulating, acting hypocritically, and lying. He compares Bolzano’s view of lying with Frege’s, which does not treat lying as a form of deception. Künne criticizes Frege’s simpler view, arguing that a liar’s “intentional profile” involves both “thematic” and “expressive” deceptive intentions. But he also criticizes Bolzano’s view, and offers his own definition, combining and improving on both Frege’s and Bolzano’s. During the course of his paper, Künne discusses a range of examples, both his own and ones drawn from philosophical discussions and works of literature. The paper as a whole illustrates very well just how subtle and complex are the interrelationships between the various concepts that are invoked in analyzing what might initially seem to be fairly straightforward mental acts such as those of deception.

(C.P.)

III. Intentionality

Intentionality is the mental capacity to represent or to refer to objects, properties or states of affairs; sometimes, in looser language, scholars speak of “aboutness”, meaning the capacity of our mental states and processes to be “about” objects and events. More specifically, scholars say that intentionality is the property that makes a mental state a representation, which is the property of a mental state of bearing a semantic relation to something else. The term, used in the middle ages, was revived in contemporary philosophy by Franz Brentano and his pupil

Edmund Husserl. The debate about intentionality has been developed recently, especially in the fields of philosophy of language and mind, starting with the works of Roderick Chisholm and John Searle.

In his criticism of Artificial Intelligence Searle (1980) suggested that the main difference between humans and machines is intentionality: artificial systems have a syntax, but they lack intentionality, while humans, endowed with natural brains, have intentionality. Searle's challenge provoked many reactions and compelled scholars to pay attention to a bunch of problems that had been largely left at the margin of the discussion because of the opposition and distrust engendered by Quine to both intensional and intentional entities. In the papers included in this section we will see three points of view, a naturalist one, an anti-naturalist one and an intermediate one, dealing with the idea of collective intentionality. We might list the problems as the three conditional questions:

- (i) *naturalization of intentionality*: if intentionality is something natural (Searle says that only beings with brains may have it), is it possible to reduce it to a purely physical phenomenon? And if so, why could artificial entities not have it?
- (ii) *intentionality and norms of thinking*: if we wanted to support the traditional account of intentionality, how can we avoid falling into the deep waters of spiritualism and idealism?
- (iii) *collective intentionality*: if intentionality is the capacity to refer to objects and events, is it possible to postulate collective intentionality beyond an individual one?

(i) Naturalisation of intentionality

The naturalisation of intentionality, as SANDRO NANNINI says in "Intentionality Naturalised", comes in two forms, one weak and one strong. Weak naturalisation provides a functional reduction of mental states: mental states are reduced to functional states that, given certain inputs, yield certain outputs. Strong naturalisation formulates hypotheses as to how the functional states are implemented in the brain.

Following this perspective, functionalism would be the first step towards the naturalization of intentionality: we need therefore to have at least a clear idea of the reduction of mental states (belief, desire, and so on) to functional states. Nannini considers, first, the view of identifying the functional role of mental or intentional states with causal co-variation.

According to this view, a cognitive system has an intentional state that is a representation of x if and only if it has an internal state y that co-varies with the presence of x , as a matter of nomological necessity. As it stands, this view is affected by a difficulty. Suppose a state y co-varies with the presence of horses: the presence of horses causes nomologically the occurrence of the state y in the cognitive system. One would be justified in concluding that y is a representation of horses. (Cf. Fodor, 1987) The problem is that there might be nomological co-variations between the occurrence of the state y and the presence of things other than horses. There might be certain conditions, say, in darkness and from far away, in which the presence of cows might cause the occurrence of the state y as a matter of nomological necessity as well. If that were the case, then the state y would be a representation of everything that is a horse or a cow seen in darkness and from far away. So the problem arises of distinguishing the correct applications of the representation y to horses from the misapplications of it to cows seen in darkness and from far away.

One reply to this difficulty consists in selecting those nomological co-variations that take place in optimal conditions. Optimal conditions are individuated as those conditions where any application of the representation y is guaranteed to be correct. This proposal encounters an objection of circularity. Optimal conditions must be specified in such a way as to prevent the possibility of misapplications and employing only naturalistic notions without resorting to intentional terms. Doubts have been cast on the possibility of characterising optimal conditions in non-intentional terms. For example, it has been argued that at least the definition of optimal conditions should involve a specification of those conditions characterised by the absence of any belief that might mediate the occurrence of the state y from the presence of things that are not x . The pressing problem is that the notion of belief is an intentional notion.

The solution advocated by Nannini is a version of the teleological reduction of intentionality. According to the teleological view, the optimal conditions are those conditions that are evolutionary significant (Millikan, 1984). Optimal conditions are those conditions in which the co-variation serves to perform a function that is evolutionary or biologically useful. For example, a co-variation between the occurrence of certain states and the presence of things of a certain sort might engender the ability to discriminate food or predators and serve to perform certain actions like eating or escaping. In general, then, Nannini suggests taking into account not only causal relations between the presence of things and the occurrences of internal states, but also causal relations between the occurrences of internal states and the performance of certain functions.

Nannini concludes that the programme of giving a teleological reduction of intentionality might be successfully pursued and might be also the first step toward the naturalisation of other mental phenomena. The naturalisation of consciousness and the self will be the next steps. The paper is programmatic, but gives a reconstruction of a typical discussion of the topic.

(ii) intentionality and objective content

One of the main problems for a naturalistic reduction of intentionality is the problem of non-denoting terms. How can our mental processes be about something which does not exist? In “Thought and Thinking: the Ontological Ground of Intentionality”, ELISABETTA SACCHI outlines what she calls “the traditional picture of intentionality”, and elaborates on what she takes as its most defensible form. On this picture, a mental act is *about* an object or objects in virtue of that act having an appropriate *content*. Such a picture, she writes, involves the following two claims about contents:

- (i) contents are some kind of entities figuring as immediate relata of mental acts;
- (ii) contents have the functional role of directing our mental acts onto what they are about.

As it is apparent in the paper of Sacchi, contents are equated to Fregean thoughts understood in the tradition of John McDowell’s conception of *de re* thoughts, (that is thoughts which are dependent on the existence of the objects they are about). Therefore, with regard to (i), she argues that contents must be seen not as self-subsistent entities (whether Platonic or psychological) but as ontologically dependent entities – entities dependent for their existence on both the mind and the world. The suggestion is that “a given act of thinking has the property of being about an object O in virtue of having a content which depends for its very existence on that object” (p. 144). We have here an original use of the Fregean distinction between thinking as mental process and thought as the content of such a process that we have defined in the first part of this introduction.

In the remainder of her paper, Sacchi discusses various problems with this suggestion and possible responses to them. She first notes that such ontological dependence is clearly too weak a requirement for intentionality, and proposes adding the requirement that “the bearer of the act stand in a cognitive relation with the object the act depends on for its very existence” (p. 146). But this modified version, too, is open to objection, and in particular, faces the problem of non-existent entities.

What we have, in fact, is a conflict between claims (i) and (ii) above and what Sacchi calls “Brentano’s requirement” – that any adequate account of intentionality must allow that we can have thoughts about non-existent entities (cf. p. 140). She considers how such a conflict might be resolved, and in the conclusion suggests that we may need to broaden our view of what can count as the *relata* of an intentional relation (to allow non-existent entities). She does not advocate any particular resolution, but she does remark that there may be room to develop a non-naturalistic theory.

(iii) collective intentionality

After having displaced the centre of his interest from the philosophy of language (*Speech Acts*) to the philosophy of mind (*Intentionality*), John Searle has devoted much of his efforts to the treatment of the social aspect of language and mind. The link between his former views with the latter is given by the idea of collective intentionality. Searle (1995), in fact, proposes an account of complex social phenomena with a limited number of conceptual tools, including the concept of intentionality. The first problem is how to treat the intentionality of social agents: is intentionality a mental property typical only of individuals or can it be seen as a property of collective entities? Searle takes the latter choice at the core of his philosophical view, recognizing collective intentionality as a peculiar primitive phenomenon: collective social entities, such as groups and other social institutions, have their own capacity to refer to objects, properties or states of affairs.

In “Is Collective Intentionality Really Primitive?” ELISABETH PACHERIE challenges some of Searle’s assumptions, partly relying on Michael Bratman (1999). As we have seen, Searle treats collective intentionality as a biologically primitive phenomenon, irreducible to individual intentionality. The same social performance may be the result of chance (different individual intentions) or of collective intentions (think of people moving at the same time because of individual reactions about some happening or because they have a common goal, as in a dance performance). Searle wants to explain the difference, contrasting an action provoked by many individual occurrences of “I intend” with an action provoked by a collective “we intend”. Pacherie points out two different claims in Searle’s explanation: (i) collective intentions are different from the sum of individual intentions; (ii) collective intentions cannot be analysed in term of individual intentions. While accepting (i), Pacherie criticises (ii), showing some inconsistencies in Searle’s view, particularly in his attempt to make collective intentionality compatible with the constraints of an internalist point of view: “it is one thing to claim that all

the intentionality there is, including collective intentionality, is in the head of individuals. It is another to insist that collective intentionality could be had by a single individual, not to mention a brain in a vat. The first claim is at least plausible, the second seems to me unintelligible” (p. 163).

The alternative proposed by Bratman does not postulate any mysterious form of primitive collective intentionality plugged into the brain of the agents, but tries to capture what is distinctive of collective intentions in terms of a special kind of interdependence of the individual intentions. Bratman analyses the grounds of shared cooperative activity, where we need to have shared intentions. Being shared, collective intentions are characterized as necessarily relational. In this way collective intentions require mutual explicitness of intentions and plans, which requires in its turn linguistic communication. Given that, Bratman’s analysis works well for behaviours that involve rational deliberation and conscious planning, but it cannot explain many cases of apparent collective intentionality in action. In conclusion, Searle’s account is shown to be question-begging and Bratman’s account is shown to over-intellectualize collective intentionality. Although rejecting the positing of a *sui generis* kind of primitive psychological attitude, Pacherie does not abandon Searle’s overall project and tries to mix it with a weakening of Bratman’s view of shared intentions. Relying on a large amount of neurophysiological results the author suggests abandoning the requirement of mutual *knowledge* and replacing it with the weaker requirement of mutual *presumption* - that is the implicit assumption that other human agents are cognitively similar to us and we can simulate or infer their attitudes or intentions. There is no a priori guarantee that our presumptions work, but they are warranted by the satisfaction of the mutual expectations of the agents. Although this stance appears to be strictly externalist, Pacherie argues that, distinguishing between “actual” and “would be” intentions, the theory requires externalism for “actual” intentions but does not supply any grounds for rejecting internalism for “would be” intentions. (C.P., M.V.)

IV. Naturalism

Whether and how the mental finds its place in nature are questions that naturalists try to answer. The papers collected in Part IV address topics concerning naturalism and naturalistic approaches to the mental. It is commonly agreed that naturalism is more a general philosophical attitude than a theory or kind of theory. The root of naturalism is the empiricist view that our cognitive capacities must be explained in accord with the

idea of ourselves as physical organisms that bear causal relations to the environment. This is the view that cognitive capacities are grounded in the architecture of our cognitive system and in causal relations to the world.

The empiricist view has influenced two basic strands within naturalism: one has an ontological orientation and the other has an epistemological one. From the ontological standpoint, naturalism implies the rejection of whatever supersedes the natural. From the epistemological standpoint, naturalism implies the rejection of the view that conceptual analysis is a source of a priori and unrevisable knowledge. As MARIO DE CARO points out in his paper included in Part Four, the conjunction of the ontological and the epistemological strands gives rise to three core theses of naturalism:

- (1) *The Constitutive Thesis*: philosophy should not appeal to any supernatural concept.
- (2) *The Anti-foundationalist Thesis*: philosophy does not have any foundational role in respect to science.
- (3) *The Continuity Thesis*: philosophical investigations are to be continuous with science.

Thesis (1) captures the idea that any explanation that appeals to supernatural entities, properties, events and capacities is illegitimate because it purports to resort to something that does not exist. For example, following G.E. Moore, one might hold that ethical judgements are about the supersensible property of being good. The property of being good would be inaccessible through the senses and undetectable in any causal interaction. It needs to be explained, then, why certain actions are good by postulating the presence of the mysterious property of being good, whose detection should be explained by postulating a mysterious mental capacity of a special intuition, totally detached from senses and the cognitive architecture of our brain.

On the other hand, theses (2) and (3) follow from the idea that conceptual analysis, if any, cannot be the source of a priori and unrevisable knowledge. Foundationalism is the view that philosophy exploits conceptual analysis to pursuing the goal of justifying scientific, epistemic and methodological principles on unrevisable bases and from outside science. But if conceptual analysis is not credited with the capacity to engender unrevisable knowledge, as naturalists claim, then foundationalism needs to be rejected. Furthermore, no privileged and a priori perspective from which our scientific, epistemic and methodological principle can be justified remains in place, apart from the standard

scientific canons of predictive adequacy and simplicity, which, however, do not guarantee unrevisability.

In what follows, in giving the summaries of the papers included in this Part, we will discuss examples of the ontologically oriented naturalistic position, the epistemically oriented one and the “liberal” one.

(i) The ontological point of view

The general empiricist picture that characterises naturalism influences the ontological view that everything that exists is an empirically accessible part of nature. The crucial point to understand naturalism, then, is to say what counts as an empirically accessible part of nature. As there are different views on what counts as an empirically accessible part of nature, naturalism assumes different versions and some are more restrictive than others.

Some philosophers define what counts as an empirically accessible part of nature by privileging the ontological strand in naturalism. We have a cognitive access only to things to which we bear causal relations and we bear causal relations only to things that belong to the physical world. The physicalistic metaphysics assumes the principle of the *causal closure of the physical*: every physical event has a sufficient physical cause, if it is caused at all. No non-physical entities (properties, states, events) are needed to explain the world we inhabit and our knowledge of it. The physical world is empirically accessible through impingements on sense organs and nature coincides with it. This view, then, characterises naturalism as primarily an ontological doctrine. Everything exists in the causal order of nature: if anything exists, either it belongs to the physical world or can be reduced to it. Conversely, if something is not physical or is not reducible to the physical order, it must be eliminated from our ontology as a fiction.

There are two kinds of theories stemming from the ontological stream of naturalism: reductive physicalism (or type identity physicalism) and eliminative physicalism.

Type identity physicalists (Place, 1956; Smart, 1959) hold that mental states (events, properties) are identical to physical states (events, properties) to the extent that any two agents are in the same (type of) mental state if and only if they are in the same (type of) physical state. Physicalists of this kind call for definitional reductions that establish type-type correlations between mental states (events, properties) and physical states (events, properties). Accordingly, mental states owe their identity to their physical constitution. What makes something a mental state is its physical structure. For example, the mental state of being in pain ought to

be identified with the state of having certain fibres stimulated (Place, Smart and Armstrong). Type identity physicalism implies the counterintuitive idea that mental states cannot be realised in organisms with a physical constitution different from that of human beings.

Eliminativists (Quine, 1960, 1969; Churchland P.M., 1981; Stich, 1983; Churchland P.S., 1986) hold that mentalistic concepts are irremediably mistaken and must be expunged from the scientific description of the world, just because mentalistic concepts are recalcitrant to definitional reductions. Mentalistic notions ought to be regarded as similar to the notion of *phlogiston*. Natural sciences will prove that we must do without them. Thus, whereas type identity physicalists hold that mental terms refer rigidly to physical states, eliminativists claim that mental terms are empty. Therefore, alleged descriptions that make use of mentalistic terms are not real descriptions; they do not speak of anything. Of course we all make use of mentalistic concepts, and eliminativists do not recommend that we stop. However, they draw a divide between those conceptual frameworks, the scientific ones, that can be used to represent how the world is and what the world contains, and those conceptual frameworks, like the mentalistic vocabulary of folk psychology, that, while having an indispensable pragmatic or heuristic significance, are not fact-stating (Dennett, 1987). At the end of the development of scientific investigation, neuroscience only will be credited with a representative content (Quine, 1969; Churchland, P.S., 1986).

In “Do Concepts exist? A Naturalistic Point of View”, MARCELLO FRIXIONE takes an eliminative stance towards the notion of a concept. Frixione questions the thesis that concepts form a natural kind. From the point of view of cognitive sciences concepts are to be identified with localised structure in the architecture of the cognitive system. Frixione casts doubts on the view that such identification can be worked out. Apart from the negative results that prove the failure of the classical theory of concepts according to which concepts are identified with definitional structures, there is no unanimous agreement on the interpretation of the existence of prototypical effects in categorisation. The existence of homogeneous structures, that is prototypes, is only one among different hypotheses that have been proposed to explain the prototypical effects.

Frixione holds that the prototypical effects might be explained by ecological constraints that the environment puts on any cognitive system evolving and interacting with it rather than postulating the existence of homogeneous structures underlying categorisation. He also mentions recent results in cognitive sciences that seem to favour theories that combine aspects of the prototype view, of the exemplar view and of the

knowledge approach. However, even this hybrid approach faces some difficulties. First of all some empirical criterion is needed in order to single out conceptual knowledge from factual knowledge. Moreover, the same piece of knowledge might be relevant for the individuation of several concepts. It seems that, even assuming a principled distinction between conceptual knowledge and factual knowledge, one is hardly guaranteed that the problem of determining the concepts that are singled out by each piece of knowledge allows for clear solution. Frixione argues that not even the neural localization of certain categorisation abilities speaks directly in favour of the individuation of conceptual systems. His conclusion, then, is sceptical: from the standpoint of cognitive science there are good reasons for doubting the existence of concepts.

(ii) The epistemological point of view

Other naturalist philosophers try to define what is an empirically accessible part of nature by privileging the epistemological strand in naturalism. In general, from this perspective everything is an empirically accessible part of nature that is required or allowed by the ontology and the methodology of natural sciences. Sorts of entities (properties, event, states) that lack causal efficacy might nonetheless be counted as empirically accessible if the theories that quantify over them form indispensable parts of the body of natural sciences that, as a whole, has an empirical content. The typical example of this approach is Willard Van Orman Quine's attitude towards mathematical entities. Mathematics is a necessary part of natural science. Therefore, if we are to treat scientific theories as true, given that mathematics is part of them, mathematics too is to be taken as true. But mathematics can be taken as true only if the entities in its domain of quantification exist.

There is a kind of physicalism that does not promote ontological reduction (in the sense of type identities) of the mental to the physical. This kind of physicalism stems from the aforementioned epistemological strand in naturalism. Accordingly, the mental is not identical to the physical; the mental supervenes on the physical at least in the weak sense that there are no changes in the mental without changes in the physical. Functionalism (Fodor, 1968; Putnam, 1975) is a version of this kind of physicalism. Functionalism holds that the identity of mental states is determined by the functional relations they bear to the agents' sensory stimuli, behavioural outputs and other mental states so characterised. Given that sensory stimuli and behavioural responses are allowed by the ontology and methodology of natural sciences, so can functional states.

The advantage of functionalism is that it makes it possible for the same

mental state, identified as a functional state, to be realised in different ways in organisms with different physical constitutions. In other words, mental states can be multiply realised. For example the mental state of being in pain can be functionally defined as that state such that (i) is caused by injury, (ii) causes the belief that one is within a situation to be avoided, (iii) causes the desire to put an end to that situation, (iv) causes worry and moans. It might be the case that the occurrences of the state of being in pain are nomologically correlated to the occurrences of a certain neural state in organism of a certain kind. But the mental state of being in pain does not owe its identity to such nomological correlation.

Functionalism, however, has been charged with instability (Kim 1993, 1998). Indeed, if functionalism allows for an account of the manner in which functional states are implemented into the physical structures of the agents, then functionalism implies explanatory reductionism, that is, the derivation of psychological laws and truths from the laws of some underlying physical theory together with “bridge” laws correlating mental states functionally defined with physical states. Bridge laws establish nomological correlations between mental states and physical states. These nomological correlations support strong supervenience: necessarily, for any mental property *M*, if an agent *a* has *M*, then there is a physical property *P* such that *a* has *P*, and necessarily if anything has *P*, it has *M* (here “necessarily” stands for nomological necessity). If one takes the union of the properties *P*s with which a given mental property is nomologically correlated, one gets a reduction base for *M*. The union of *P*s is a complex disjunctive property that is necessarily (in the nomological sense) coextensive with the mental property *M*. In the end, one obtains a reduction of the mental property *M*.

On the other hand, if functionalism denies the possibility of an account of the manner in which mental properties are implemented into the physical structures of the agents, then it cannot avoid the possibility that even weak supervenience is violated. And some philosophers hold that a view that does not make the dependence of the mental upon the physical robust enough to support at least weak supervenience can hardly be counted as physicalist.

In “Cosmic Hermeneutics vs. Emergence: the Challenge of the Explanatory Gap”, TIM CRANE argues that the problem of the explanatory gap gives rise to a concern about the proper formulation of non-reductive physicalism. The concern is about the manner in which non-reductive physicalism has to be formulated in order not to collapse into emergentism. The explanatory gap consists in the fact that we lack an adequate explanation of all the truths about consciousness (and the mental

in general) in physical terms. It is assumed that the explanation must be in the nomological-deductive form according to which an explanation is a deductive argument that derives the *explanandum* from the *explanans*. Physicalism entails global supervenience: every minimal physical duplicate of the actual world is a duplicate in every respect. Global supervenience is a necessary condition for physicalism but not a sufficient one. Indeed, global supervenience is endorsed by emergentism as well. Emergentists hold that mental properties (i) cannot be reduced by reductive definition to physical properties, (ii) cannot be explained from physical properties, and (iii) cannot be “added from outside” as the Cartesian conception of the mental claims. What distinguishes physicalism from emergentism, then, is the stronger thesis that the physical has an ontological priority, as reductive physicalists claim, or an explanatory priority. The thesis of ontological priority entails that all states belong to a subclass of the class of physical states. The thesis of explanatory priority entails that all truths can be explained in principle in broadly physical terms. While accepting the global supervenience thesis, emergentism is not committed to the ontological priority nor to the explanatory priority of the physical over the mental. Emergentism assumes global supervenience to be a brute metaphysical fact, which we are not able to explain. The problem for those physicalists who reject both ontological reduction and explanatory reduction should be now clear. Non-reductive physicalism aims to differentiate itself both from reductive physicalism and emergentism. Doubtless, non-reductive physicalism renounces the ontological priority of the mental over the physical. So non-reductive physicalism seems to be forced to accept the explanatory priority thesis, otherwise it would collapse into emergentism. This result is puzzling for those philosophers who hold that physicalism can recover from the problem of the explanatory gap simply by rejecting the view that physicalism is committed to giving an explanation of the mental in physical terms. There is, they say, no metaphysical gap, so physicalism is true, although we cannot fully understand it just because of the explanatory gap. Crane’s analysis shows that the explanatory gap leaves physicalism with two options only. Either physicalists answer that the explanatory gap is of little importance because soon or later we will arrive at a satisfactory explanation. Or, granting that there cannot be any nomological-deductive explanation, physicalists must provide another sort of explanation. What physicalists cannot do is to claim that there is no need to give an explanation at all. If they do so, physicalism becomes indistinguishable from emergentism.

In “What Physicalism Should Provide Us With”, ACHIM STEPHAN and