

FRENCH STUDIES IN THE PHILOSOPHY OF SCIENCE

BOSTON STUDIES IN THE PHILOSOPHY OF SCIENCE

VOLUME 276

Editors

ROBERT S. COHEN, *Boston University*
JÜRGEN RENN, *Max Planck Institute for the History of Science*
KOSTAS GAVROGLU, *University of Athens*

Editorial Advisory Board

THOMAS F. GLICK, *Boston University*
ADOLF GRÜNBAUM, *University of Pittsburgh*
SYLVAN S. SCHWEBER, *Brandeis University*
JOHN J. STACHEL, *Boston University*
MARX W. WARTOFSKY†, (*Editor 1960–1997*)

For other titles published in this series,
go to www.springer.com/series/5710

FRENCH STUDIES IN THE PHILOSOPHY OF SCIENCE

Contemporary Research in France

Edited by

ANASTASIOS BRENNER

Université Paul Valéry - Montpellier III, France

and

JEAN GAYON

Université Paris I - Panthéon Sorbonne, Paris, France

 Springer

Editors

Anastasios Brenner
Département de philosophie
Université Paul Valéry-Montpellier III
route de Mende
34199 Montpellier cedex 5, France
anastasios.brenner@wanadoo.fr

Jean Gayon
Institut d'histoire et de philosophie
des sciences et des techniques
Université Paris I-Panthéon Sorbonne
13 rue du Four
75006 Paris, France
gayon@noos.fr

ISBN 978-1-4020-9367-8

e-ISBN 978-1-4020-9368-5

Library of Congress Control Number: 2008944077

© 2009 Springer Science+Business Media B.V.

No part of this work may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, microfilming, recording or otherwise, without written permission from the Publisher, with the exception of any material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work.

Printed on acid-free paper

9 8 7 6 5 4 3 2 1

springer.com

Contents

Introduction

Anastasios Brenner and Jean Gayon	01
---	----

Part I: Styles in Philosophy of Science and Technology

1 Anne Fagot-Largeault: The Legend of Philosophy's Striptease: Trends in Philosophy of Science	25
2 Daniel Parrochia: French Philosophy of Technology	51

Part II: General Philosophy of Science

3 Anastasios Brenner : A Problem in General Philosophy of Science: The Rational Criteria of Choice	73
4 Sandra Laugier: Science and Realism: The Legacy of Duhem and Meyerson in Contemporary American Philosophy of Science	91

Part III: Physical and Chemical Sciences

5 Gilles Cohen-Tannoudji: Philosophy and 20th Century Physics	115
6 Hervé Zwirn: Foundations of Physics: The Empirical Blindness	141
7 Bernadette Bensaude-Vincent: Philosophy of Chemistry	165

Part IV: Life Sciences

8 François Dagognet: Pharmacology as a Philosophical Object	189
9 Jean Gayon: Philosophy of Biology: An Historico-critical Characterization	201
10 Claude Debru: Philosophy and Contemporary Biological Research	213

Part V: Philosophy of the Behavioral and Cognitive Sciences

11 Joëlle Proust: What Is a Mental Function? 227

12 Daniel Andler: Philosophy of Cognitive Science 255

Part VI: Philosophy of Economics

13 Philippe Mongin: Duhemian Themes in Expected Utility Theory 303

Notes on the Authors 359

Name Index 367

Introduction

Anastasios Brenner and Jean Gayon

The first two sections of this introduction provide an overview of the history of philosophy of science in France. The last section comments on the specific content and structure of the present book.

1 An Attempt at Periodization

For those who first encounter French philosophy of science the prominence of Gaston Bachelard (1884–1962) comes as somewhat of a surprise. His thought does not fit in easily with the familiar currents: neo-Kantianism, pragmatism, logical positivism. His own labels – applied rationalism and technical materialism – are perplexing. The reception of his works in English-speaking countries was belated, and his influence there has been slight. Yet Bachelard was a prolific and an inventive author. Trained in mathematics, physics and chemistry, he was quick to turn philosophical reflection toward the new theories of his day: atomism, relativity and quanta. His relentless efforts yielded some 20 books. From philosophy of science Bachelard was led further to the study of imagination and poetry. He did not fail to conceptualize his experience as a science teacher and later a professor of philosophy, providing incentive remarks that could inspire pedagogical reform. His

A. Brenner (✉)
Département de philosophie
Université Paul Valéry-Montpellier III
route de Mende
34199 Montpellier cedex 5, France
E-mail: anastasios.brenner@wanadoo.fr

J. Gayon
Institut d'histoire et de philosophie
des sciences et des techniques
Université Paris I-Panthéon Sorbonne
13 rue du Four
75006 Paris, France
E-mail: gayon@noos.fr

observations on the role of science in society could even be called upon in the context of political thought. Georges Canguilhem (1904–1995), claiming to pursue his legacy in the philosophy of medicine, initiated what can be termed a Bachelardian school. In turn Michel Foucault (1926–1984) brought this school to the attention of an international audience. Bridging at least three generations, Bachelard's thought continues to be felt in the works of several authors in this volume.

Yet when Bachelard came on stage, philosophy of science was already a well-established discipline in France. His two thesis supervisors, Léon Brunschvicg (1869–1944) and Abel Rey (1873–1940), had provided him with methods of inquiry as well as an institutional setting. The former had initiated him to an original brand of neo-Kantianism and convinced him of the decisive character of mathematical reasoning in the sciences. Abel Rey had shown him how to elaborate a precise philosophical reflection with respect to earlier changes in science: thermodynamics, electromagnetism and atomic theory. There are also those important figures of the past generation whose ideas at first inspired Bachelard and then to which he reacted: Henri Poincaré (1854–1912) and Pierre Duhem (1861–1916). Finally, one should not forget the importance of Émile Meyerson (1859–1933) with whom Bachelard was to constantly measure himself.

To reach the inception of philosophy of science, we must go back still further. Auguste Comte (1798–1857) is certainly to be counted as one of its founding fathers.¹ His *Cours de philosophie positive* provided an impressive picture of the complete spectrum of the sciences and initiated several major topics of this new field of studies.² The 50-year period that begins with the publication of the first volume of this work in 1830 and ends in 1870 we may take as the founding years. Various historical factors can be invoked to explain the emergence of philosophy of science. The chemical revolution, the impact of industrial developments showed the need to reconsider the synthesis provided by Diderot and d'Alembert in their *Encyclopédie*. The French revolution carried with it projects of social reorganization and educational reform. Comte's initiative runs parallel to attempts occurring in Europe at the same time, provoked by similar causes but carried out in different manners: for example Bolzano's *Wissenschaftslehre* (1837) and Whewell's *Philosophy of the Inductive Sciences* (1840).³

Comte set the agenda in several respects for philosophy of science in France. Positivism, in one form or another, dominated here the philosophical scene until World War One, and even later thinkers who had relinquished positivism continued to pay tribute to him, most notably Canguilhem and Michel Serres.⁴ First and foremost there is Comte's decision to favor a historical approach over a logical one. Philosophy of science, he continually asserts, must be grounded on history of science. This trend was to characterize in the main French philosophy of science. As an attempt to direct philosophical reflection toward science and to make scientific knowledge a model, positivism, in its various forms, has been intimately bound up with a large portion of philosophy of science either as a source inspiration or as a target for criticism: from Comtian positivism to logical positivism and even to post-positivism. It is thus important to come to grips with the significance and role of this doctrine. The Vienna Circle acknowledged a debt toward Comte, and one finds in the latter at times some astonishing anticipations: Comte formulates an empirical

criterion of meaning in order to exclude metaphysics and adopts with respect to the development of science a resolutely sociological approach.⁵

Some of the questions considered essential by Comte however lost their importance or disappeared altogether. He went to great pains to draw up a classification of the sciences and delimit the fundamental disciplines. By the end of the 19th century science had given rise to much interdisciplinary research, and readers of Comte could ironize on the limitations he imposed. On the other hand, topics to become central such as measurement or theory confirmation did not receive much attention in his works.

There were other important figures from this period: André-Marie Ampère (1775–1836), Antoine-Augustin Cournot (1801–1877) and Claude Bernard (1813–1878). Ampère's *Essai sur la philosophie des sciences* (1834) is among the early works to signal an autonomous reflection on science. Although it quickly fell into oblivion, it is part of the story, as a book with respect to which both Comte and Whewell could contrast their own conceptions.⁶ Ampère provides a link with the philosophizing that went on in the vigorous scientific movement of the 18th century as well as a close reworking of the conceptions of the encyclopedists. He is however best remembered for the controversial claim he made with respect to his major scientific contribution that his mathematical theory of electrodynamic phenomena was "entirely deduced from experience". More subtle perhaps is the emphasis he places, in introducing his theory, on the descriptive task of science and its separation from metaphysical speculation. This prepared the ground for phenomenalist conceptions. It also showed the need to develop a discipline that would take up those questions left unanswered by science.

Cournot produced a series of noteworthy philosophical works, associating subtle philosophical analysis with careful historical study, among others *An Essay on the Foundations of Our Knowledge* (1851). In these works he breaks with the limits imposed on science by Kant and goes beyond Comte's harnessing of science for social purposes. He does not shrink from addressing questions of origin, that of the universe or that of life. Cournot provides an attempt to follow up the metaphysical implications of science. Thus, in his view, science leads us to conceive our solar system like an island in a boundless, uncentered universe. Cournot also develops, in contradistinction to positivism and phenomenism, a realist position in philosophy of science.

Bernard, who made a number of significant contributions to medicine, went on to reflect on his practice as a scientist. He explored the philosophical problems raised by the life sciences and brought attention to the precise features of experimental method. His *Introduction to the Study of Experimental Medicine* provided the background for discussions throughout the 19th century.

The Franco-Prussian War in 1870 not only signals a change of political system – the end of the Second Empire and the beginning of the Third Republic – but led a whole generation to reflect on French science and to seek to emulate the German university. We see here a second 50-year period running until the end of the First World War in 1918. Ushered in by Paul Tannery (1843–1904) and Émile Boutroux (1845–1921), this period is characterized by the institutionalization of the discipline. While Tannery called on Comte's pioneering studies,

he sought to go beyond his very general and at times frankly superficial survey and give rise to genuine historical research, reading and editing manuscripts and correspondences. He thereby set the groundwork for the studies of Duhem and Gaston Milhaud (1858–1918) who were intent on bringing out the philosophical implications of scientific evolution. Boutroux, although primarily a historian of philosophy, directed attention to science. Along with Charles Renouvier (1815–1903), he brought the intricate conceptual system of Kant to bear on the scientific problems of the day. To this effect he resorted to Comte, producing a peculiar brand of rationalism and positivism.

It is telling of the attraction that philosophy of science exerted in those years to learn that Henri Bergson (1859–1941) was at first tempted to embrace this specialty. As he confides to William James in a backward glance on his formative years:

My intention was to devote myself to what was then called “philosophy of science” and, with this in mind, I set out to examine some fundamental scientific notions. It was the analysis of time as it occurs in mechanics or physics that overturned all my ideas.⁷

The elaboration of his metaphysics is thus preceded by a period of philosophical scrutiny of science. And his mature thought develops claims with respect to space, time, consciousness and life that have implications for philosophy of science. There were those who sought to follow up the consequences of his thought in this direction. But through the controversies he spurred and his insistence on questions avoided by scientific research, Bergson was a thinker philosophers of science had to contend with.

A succession of scientific revolutions was to provoke a reworking of the picture of knowledge: non-Euclidian geometry, the theory of evolution, thermodynamics and electromagnetism. One of the leading figures was Poincaré. His research in mathematics convinced him that non-Euclidian geometry was not a mere fiction but a fruitful conceptual construction. Meditating on the nature of geometrical hypotheses, Poincaré advanced the idea that they are conventions. Thereby he rejected both John Stuart Mill’s claim that they are empirical facts and Kant’s contention that they are synthetic *a priori* propositions.

Duhem formulated a similar idea with respect to physics. Hypotheses are not directly derived from experience; they are founded on the free choice of the theoretician. Indeed, Duhem condemns the Newtonian method of inductions and the procedure of crucial experiment. Experimental refutation is more complex than it was generally believed. This led to the holist thesis Quine was to formulate in the context of a logical analysis of science.

These striking results were taken up by several philosophers and scientists. Édouard Le Roy (1870–1954), a mathematician who had studied under Poincaré before embracing Bergson’s philosophical perspective, perceived here the rise of an intellectual movement that he labeled – it is worth noting – “a new positivism”.⁸ This reformulation of positivism captures in part what Duhem, Milhaud and Abel Rey were striving to do. It attracted the attention of young Austrian scholars who were to found the Vienna Circle and provides a connection between the philosophical traditions of France and Austria that deserves to be emphasized. However, the aim here was to moderate Comte’s strictures on theoretical speculation and metaphysics.

Le Roy emphasized the novelty of these reflections on science; he was one of the first to make use of the term *épistémologie* or epistemology, as we shall simply transpose it here.⁹ The term designates in French usage philosophy of science rather than theory of knowledge.¹⁰ What was being proposed was an investigation precisely centered on scientific activity. This carried an implicit criticism of earlier philosophy of science, as practiced by Comte and Ampère, and signaled a shift in the discipline.

In connection with these debates over the nature of scientific theories were early ventures to introduce philosophy of science into the university curriculum. One of the first seems to have been Arthur Hannequin (1856–1905), who around 1890 proposed courses under this heading at the Faculty of letters of Lyon. In 1892 a chair of “General history of science” was instituted at the College de France. Although historical in orientation, the generality aimed at expresses a philosophical concern.¹¹ Shortly thereafter Milhaud set up a program of study in philosophy of science at Montpellier. His endeavor was crowned by success, and a chair of history of philosophy in its relation to science was created for him at the Sorbonne in 1909. This chair was to play a pivotal role in the future of the field, being held successively by Abel Rey, Bachelard and Canguilhem.

Although Meyerson belonged to the same generation as Poincaré and Duhem, he published his first book *Identity and Reality* in 1908, and his philosophy came to attract attention especially after World War One. The interwar years represent a turn in philosophical reflection on science. Several prominent figures had disappeared during the war, and the new scientific theories called for a reexamination of past views. Meyerson refused Poincaré’s recourse to induction and probability as a way of justifying hypotheses as well as Duhem’s deductive and holistic solution. He called on common reasoning and a psychological study of science. The hypotheses of science are plausible. The mind seeks identities, to which the real opposes resistance. Trained as a chemist under Bunsen and well-versed in German philosophy, Meyerson brought a different outlook as he became involved in discussions in France.

Taking up Poincaré’s ideas, Abel Rey was careful to emphasize the tendency toward realism. He was in particular struck by the recent discoveries of atomic theory. Philosophy of science was to be based on history of science, and Abel Rey was led to elaborate a historical approach based on techniques developed in the social sciences. He was fortunate enough to have his thesis *Physical Theory according to Contemporary Physicists*, a synthetic presentation of the turn-of-the-century debates, seized upon by the logical positivists. Abel Rey was furthermore included among Neurath’s collaborators to the *Encyclopedia of Unified Science*. However, this promising connection between French conventionalism and Austrian positivism was cut short. The First Congress of Scientific Philosophy was indeed held in Paris in 1935, but logical positivism presented at this early stage a radical program,¹² and French philosophers by then were trying to shake off the long influence of positivism. Whereas the Vienna Circle and those associated with it were to have a strong impact in English-speaking countries, it is to be noted that French philosophy of science followed a very different pattern of development.

Abel Rey had taken up Milhaud's chair at the Sorbonne,¹³ and he brought about the founding of the *Institut d'histoire des sciences et des techniques*.¹⁴ The aim of this institute was to encourage cooperation between the sciences and the humanities. Several prominent scientists were involved, and it gave rise to significant international collaboration. Thereby was achieved a strong institutional recognition of history of science and philosophy of science.

We may now return to Bachelard, who succeeded to Rey. He can be credited with having forcefully directed philosophical attention to the latest scientific theories. Along with Koyré, he was convinced that the succession of revolutions that had shaken science since the discovery of non-Euclidian geometry called for a "philosophical revolution".¹⁵ Borrowing a phrase from Reichenbach, Bachelard spoke of a "conflict of generations",¹⁶ and he was quickly led to spell out the inadequacies of the philosophical conceptions of his predecessors. In a typical passage of *The New Scientific Spirit* of 1934, Bachelard states:

One understands (...) the rejection of Poincaré's opinion concerning the supreme convenience of Euclidian geometry. This opinion seems to us more than just partially erroneous, and, reflecting on it, one derives more than a lesson of prudence in forecasting the destiny of human reason. By correcting it, one is led to a genuine reversal of value in the rational realm and one recognizes the essential role of abstract knowledge in contemporary physics.¹⁷

Although Bachelard emphasizes the component of action in knowledge, he refuses to let himself be drawn toward pragmatism. We know that Poincaré called on simplicity in order to defend the choice of Euclidian geometry as the language of physics. After the advent of the theory of general relativity, such a position could no longer be held. Bachelard not only invites us to take into account the global simplicity of the theoretical system; he suggests the importance of consistency. Science in its progress seeks to coordinate apparently divergent theories. This leads to a transformation of rationality itself. Bachelard developed his philosophy on broad lines. Following in this Duhem and Meyerson, he carried his philosophical inquiry into the field of chemistry – a science conspicuously neglected by the Vienna Circle.¹⁸

Bachelard clearly marks a transition in philosophy of science. One perceives in his first book, *Essai sur la connaissance approchée*, how the idea of an "inductive power of mathematics" took shape. One also learns how Bachelard realized from the outset the meaning he sought to give to words like reality and realism:

Knowledge, if it is not to go against its principle of continual increase, cannot be tautological. It must therefore imply, willingly or not, an unknown element in the substantive that analysis claims to resolve into predicates. Thus one can turn down the realism of notions while accepting as constant a progressive reification. This constructive ontology never reaches its term, because it corresponds rather to an action than a finding. Should the object be at one moment assimilated and rationalized as it were, erased as an obstacle, reduced by analysis to its true nature as a notion; the same constructive process will then relate it to a new irrational. Generalization in mathematics tends to absorb the fields that border on the primitive field [...]. This constructed realism will therefore set down a whole series of successive data. The elements will take on in these fields truly different existences, and it will be by an abuse of ontology that we shall forget the conditions that pertain only to these fields in order to transform them into properties truly belonging to the entities.¹⁹

Let us emphasize the expressions of “constructive ontology” and “constructive realism”, which summarize the epistemological ontology developed in the later works.

Bachelard is not content to reformulate positivism; he opposes positivism altogether. Bachelard did not embrace logical analysis. He continued the historical approach, but affected a change in method. A continuist conception of scientific evolution had dominated up until then, this most prominently in Meyerson’s version: scientific knowledge evolved from common knowledge, and each stage of science was historically linked with the precedent. A different model underlies Bachelard’s view of progress, which is marked by a series of discontinuities, both at the historical and the cognitive levels. It is accompanied by the insistence on a “recurrent reading” of history: today’s science reveals the potentialities of past science.

It was after World War Two that Bachelard, who had been elected to the Sorbonne in 1940, gave full compass to the philosophical claims set out in the mid 30s and acquired his ascendancy over the field. Yet invasion, fascism and deportations had profound consequences on philosophy and justify our marking out a new period. The years from 1945 to 1970 can be characterized as those of a rather autonomous development: a deepening of this heritage in a direction leading away from idealism and an interaction with the intense debates over existentialism and political theories.

Canguilhem, who succeeded to Bachelard in 1955 at the Sorbonne (chair of history and philosophy of science), is often closely associated with the latter as a key figure of the “French style in philosophy of science”. Both authors indeed emphasized the necessary link between history and philosophy of science, and they were equally opposed to any kind of analytical approach to philosophy of science. Furthermore, Canguilhem had a strong personal relationship with Bachelard, and an impressive institutional and intellectual network developed around the two of them. Thus the “Bachelard-and-Canguilhem” style moulded the *ethos* of most of French philosophy of science in the second half of the 20th Century. Nevertheless Canguilhem’s approach to philosophy of science is sensibly different from Bachelard’s. The two philosophers differed first by their subject matter. Whereas Bachelard had a scientific background in mathematics, physics and chemistry, Canguilhem started on a medical curriculum shortly before World War Two, which led him to defend in 1943 a MD thesis on the concepts of the normal and the pathological.²⁰ Under the title *Le normal et le pathologique*,²¹ this work remains today one of the most important in philosophy of medicine. It has been translated into English²² and in a number of other foreign languages. Though Canguilhem’s reflection on the life sciences extended far beyond medicine, it continued to be deeply influenced by his primary interest in medicine. A second important difference between Canguilhem and Bachelard is the former’s major interest and investment in history of science. For Canguilhem, the effective practice of history of science was immensely important for philosophy of science. Almost all of Canguilhem’s writings after 1943 closely combine the most rigorous approach to history of science (which earned him the prestigious Georges Sarton Medal) and a kind of philosophy of science based on the conviction that the genesis of concepts sheds irreplaceable light upon the problems

of philosophy of science.²³ This doctrine and, probably more importantly Canguilhem's vivid example, certainly contributed to shape the spirit and method of the vast majority of French philosophers of science in the past 50 years.²⁴

In the late sixties and early seventies two new directions emerged: on one hand, the application of philosophical reflection to a larger variety of objects and, on the other, an opening toward other traditions.

Foucault acknowledged his debt toward Canguilhem,²⁵ who reversely considered that Foucault had more radically accomplished his own program. He developed Bachelard's and Canguilhem's "historical epistemology",²⁶ giving it a broader and more systematic orientation. Foucault did not indeed limit himself to philosophy of science proper, but sought to track down the interrelation of our elements of knowledge. He contested traditional separations, emphasizing the relation between knowledge and power. This led to his archeology of knowledge and his historical ontology. Whereas Bachelard was interested in revolutionary science, Foucault turned toward the reorganization of systems of knowledge over the long term. He sought to map out the trajectory of concepts and practices, that is the precise conditions of their historical constitution.

There is also a close connection between Canguilhem and François Dagognet (see the chapter by this author in the present volume). After training both in philosophy and medicine (psychiatry), Dagognet (born in 1924) who was also Bachelard's *protégé*, first devoted himself to history of biology and medicine (resulting in a major book on Pasteur),²⁷ and to philosophy of medicine (his book on "remedy").²⁸ Although Dagognet came to occupy the same chair of history and philosophy of science as Bachelard and Canguilhem, and played an institutional role quite similar to Canguilhem, his writings (numbering more than 60 books) extend to many more subjects than his mentors. Not only did he write on a huge array of scientific and technological topics going far beyond biology and medicine), but also on almost all possible subjects in philosophy, including morals, politics, religion, and art. The most distinctive trait of his thinking is to have attempted systematically to reevaluate a number of traditional philosophical problems in the light of present scientific knowledge. We know of no other contemporary philosopher who could more vividly illustrate Canguilhem's dictum: "Philosophy is a reflection for which any foreign matter is good – and we hasten to add – for which any good matter should be foreign". As Bachelard, Canguilhem, and Foucault, Dagognet always felt alien to logical empirism and analytic philosophy.

On the other hand, Jules Vuillemin, Gilles-Gaston Granger and Jacques Bouveresse attracted attention to logical empiricism and analytic philosophy, leading to an attentive examination of these currents. Vuillemin (1920–2001) came from the school of historians of philosophy that had formed around Martial Gueroult. He had followed the teaching of Bachelard and Cavaillès. In his early works on Descartes and Kant he endeavored to bring out the connection between their metaphysical ideas and the scientific method which inspired them; he was intent on revealing the axiomatic organization underlying their philosophical systems. Vuillemin was thereby drawn to analytic philosophy, which he helped to propagate in France.²⁹ However he did not hesitate to express his reluctance to give up historical inquiry:

Making my way (...) a difference between me and the majority of Anglo-Saxon analysts emerged. There were those who, singlemindedly interested in chasing down grammatical errors in the talk of philosophers, forgot the existence of scientific languages. But even those who applied the method of “rational reconstruction” to these latter more often imposed on them principles of their own choice. I resisted this violence done to history, and trusted in the sciences such as they are, and not such as they should be. Moreover, it is presumptuous to neglect the philosophical tradition.³⁰

As for Granger (born in 1920), he had explicitly assigned himself from the very beginning the task of bringing together the threads of rational inquiry that had separated into two antagonist currents at the beginning of the 20th century: Anglo-American schools of thought resorting to logical and linguistic analysis as opposed to Continental ones devoted to conceptual and historical study.³¹ Bouveresse (born in 1940), in turn, devoted much energy to exploring Wittgenstein and the Austrian tradition, which he brought forcefully to the attention of the French philosophical community. Going beyond this source of inspiration he has produced an abundant series of original and vigorous studies on knowledge and science.³²

Yet when logical empiricism came to receive sustained attention in France, it was no longer a dominant movement. It had come under severe criticism both by philosophers using logical methods but rejecting its dogmas as well as by philosophers calling on history. Analytic philosophy had likewise undergone changes; logic had been in some quarters replaced by ordinary language. Thus Carnap, Reichenbach and Russell were studied along with the second Wittgenstein, Quine, Popper and Kuhn. Logical empiricism had by then become the object of critical scrutiny and historical examination. This would eventually lead to a series of studies that could be seen to converge with a program of research in the origins of logical empiricism and the sources of analytic philosophy.

2 Methods and Objects

In the course of its development philosophy of science changed considerably. Various manners of characterizing reflections on science were proposed. Although Comte was concerned with general issues, he was careful to distinguish the sciences with respect to method and object. Philosophy was thus applied in turn to each fundamental science, and Comte most often used the term philosophy with a defining adjective: “mathematical philosophy, astronomical philosophy, physical philosophy, chemical philosophy and biological philosophy”.³³ This gave rise to what has been named regional epistemologies; philosophical considerations are to be formulated with respect to specific scientific context. There is however a certain ambiguity here: is such reflection part of the sciences considered or something autonomous? In earlier usage philosophy could designate simply the essentials of a scientific field, its principles and methods. Later Ernest Renan was to coin the generic expression “scientific philosophy”.³⁴ He thereby referred to a particular tradition in French thought devoted to be sure to science but also naturalistic in orientation. In addition he was intent to promote a rigorous philosophy based on the model of science,

which could be seen to lead to a form of positivism better called scientism. The expression scientific philosophy was to remain in use alongside philosophy of science. But philosophical reflection came to mark its difference with respect to the sciences that constitute its subject matter – a metadiscourse.

What characterizes a large portion of French philosophy of science is the importance allotted to history. This is apparent in the early formulation of the discipline by Comte as well as its later institutional establishment. Of course, a historical approach can be pursued in many ways, and in the context we are examining, it indeed gives rise to a variety of styles of research. One direction consists in grounding philosophy of science upon the history of science. In the absence of empirical testing, history of science provides a means of assessing philosophical conceptions of science. This is particularly clear in Duhem. His *Aim and Structure of Physical Theory* furnished an analysis of the stages involved in the construction of a scientific theory. But this “logical analysis”,³⁵ as he termed it, was to be followed by a historical study, and the numerous volumes he devoted to the evolution of science since Antiquity bear witness to this preoccupation. Such a method was followed by many of his contemporaries, for example Meyerson and Brunschvicg. Post-positivists were later to call on this tradition in their effort to reassert the importance of history, and this was one of the trends of the French tradition that received the most sustained interest abroad. Yet if it is reasonable to require that philosophical conceptions be justified in some sense by the historical record, it is not obvious that philosophy should be modeled solely on the scientific method of empirical testing.

Another significant line of research is directed toward the history of philosophy in its relation to the sciences. It was encouraged by scientifically inclined historians of philosophy, such as Boutroux. Milhaud in introducing philosophy of science within the university curriculum was careful to link this specialty with the history of philosophy, which has always played an important role in France. The study of the scientific background provided a promising way to renew the interpretation of the great systems of the past. Koyré, in turn, formulated a philosophically oriented history of science. He revealed the philosophical motivations underlying the works of Galileo and Newton among others. The early writings of Vuillemin and Granger further scrutinized the connection of science and metaphysics in Aristotle, Descartes and Kant.

Bachelard had misgivings over antecedent conceptions of scientific growth as a continuous process, exemplified in particular by Meyerson. He set about to elaborate what has been named a “historical epistemology”.³⁶ Study of past science still retained to be sure its importance. But it was to be placed within a clearly discontinuist conception, inspired by the recent discoveries in science. Scientific revolutions are accompanied by breaks between common knowledge and scientific knowledge. Bachelard made explicit the position from which the philosopher observes the past: a reading is accomplished with respect to current science and must be continually renewed.³⁷ This perspective set the agenda for the intricate and subtle historical studies pursued by Canguilhem and Foucault.³⁸

Recently has emerged a conscious effort to understand the *history* of philosophy of science. French studies in this area are not unrelated to an ongoing program of

research at the international level.³⁹ One of the aims of this program is to explore the origins of logical empiricism and the sources of analytic philosophy. Different motivations may underline this historical perspective: one may wish to encourage a reappraisal of these movements, after the diverse criticisms leveled at them by their antagonists or hope to bridge the divide between Anglo-American and Continental traditions. This program has taken a particular turn in France leading also to a study of phenomenology, a current whose influence was strong. Husserl's connection with Austrian thought has been emphasized. In addition French philosophy of science prior to Bachelard has become the object of a more thorough and systematic investigation.⁴⁰

Philosophy of science, which during its early development had concentrated on theory, gradually expanded its compass to include the concrete aspects of scientific activity. The advent of a philosophy of action, Bergson's claim that humankind is best described as *homo faber* and turn-of-the-century debates over experimental method initiated a shift. Each scientific instrument has its own theory, whose purpose is to explain how it works and how it is to be applied. In turn the theorist must take into account the approximate nature of observations. Observatories and laboratories are vast networks of instruments. A sharp boundary between theory and practice is brought into question. Instrumentation reacts on the theoretical construction.

Nevertheless Bachelard's move away from the idealism that had dominated until then opened up a new path. The "applied rationalism" and "instructed materialism" of his later thought made it possible to truly take into consideration other aspects of science. His "phenomenotechnics" was seen as a call for a precise description and careful analysis of the material aspects of science. This led to an interest in machines and technology that his followers were to take up: Canguilhem, Gilbert Simondon (1924–1989) as well as François Dagognet and Bruno Latour (born in 1947).⁴¹ Philosophical reflection was applied to numerous unusual objects: factories, automata, computers, airplanes, spaceships. Along similar lines Dominique Lecourt (born in 1944) has explored the interaction between humankind and science, the fears provoked by technology and the myths surrounding progress.⁴²

Although there was some resistance on the part of French philosophers – and also mathematicians – to the development of mathematical logic, it would be wrong to think that there was no interest in applying formal methods in philosophical reasoning. Poincaré, who was preoccupied by the foundations of geometry and arithmetic, provided an examination of the construction of our notion of space, which calls resolutely on mathematical reasoning. Duhem, who was in the process of developing a highly deductive presentation of thermodynamics, formulated a precise analysis of the structure of physical theories. He came close to what has been called the standard view of scientific theories.⁴³ If modern logic makes it possible to reach a more precise formulation, the main outline was nevertheless in place: theories are axiomatic systems in which the theoretical concepts receive meaning in terms of observables with the help of correspondence rules. One understands how logical positivists could find inspiration in the work of these philosopher-scientists. During the first half of the 20th century there were indeed several significant attempts in applying logic or mathematical methods to philosophical problems by Louis Coutu-

rat (1868–1914), Jean Nicod (1893–1924), Jean Cavaillès (1903–1944), Jacques Herbrand (1908–1931), Albert Lautman (1908–1944), all tragically cut off short – that the loss was not easily compensated may point to the difficulty of forming specialists in this field within the French educational system.⁴⁴ It was as if everything had to be started over again after World War Two.

Following this interruption, there were several meritorious attempts to reestablish philosophically inclined work in logic.⁴⁵ Robert Blanché (1898–1975), although somewhat isolated, published many books on logic, its history and the foundations of mathematics, providing an account of new developments in many-valued systems and modalities. Roger Martin was instrumental in setting up a program of study in logic within the philosophical curriculum at the Sorbonne. The reception and diffusion of research in this area however was somewhat belated in France. Scholars interested in the analytic tradition, such as Maurice Boudot, Maurice Clavelin and Jean-Claude Pariente, still had to push in the 1970s and 1980s for a better knowledge of logic.⁴⁶ More recently there has been the work of Philippe de Rouilhan and Jacques Dubucs, who have established a noteworthy program in logic within the Institut d’histoire et de philosophie des sciences et des techniques. One may also signal a line of development in deontic logic, initiated by the Polish logician, Georges Kalinowski, who settled in France, followed up by Jean-Louis Gardies and more recently by Patrice Bailhache.

There has been a slow but steady increase of interest in logic in France, which has accompanied the advent of analytic philosophy. However as philosophy of mind has now come to replace philosophy of language, it is to be feared that this interest may slacken.⁴⁷

3 Philosophy of Science in France Today

In the two previous sections, we sought to present the origins, the development and the major tendencies of French philosophy of science on the long run. We shall now characterize the intentions and scope of the present volume.

The origin and purpose of the book deserves being explicated. Although the two editors are known to have dwelt on several figures and aspects of “French epistemology” (i.e. philosophy of science),⁴⁸ they did not want to produce another historical book. What we proposed to Springer was to publish in the Boston series of Philosophy of Science a collection of essays by significant contemporary French figures in philosophy of science. For several decades, the publisher had been willing to provide such a volume, but all attempts seem to have failed because of the rivalry between schools of thought and networks, and of the sometimes highly conflictual character of the relationships between individuals. The absence of a society of philosophy of science was probably both a sign of the difficulty and an obstacle to the realization of a volume of the present sort. Whatever the causes, the fact is that the project of the present volume emerged in 2004, shortly after the creation of the *Société de philosophie des sciences* in 2002.⁴⁹ The Society was not involved

in any way in the elaboration of the volume, but the context it gave rise to certainly favored the feasibility of the project through a tight cooperation between the two of us and Springer.

Here is the method that we applied. We first made a confidential list of 16 possible authors, which was raised to nineteen, the directors of the Boston series, Jürgen Renn and Kostas Gavroglu, expressing the wish that “all ideas and themes of French Philosophy of Science be represented”. Throughout the process we kept several concerns in mind. First, the length of the chapters should be reasonable enough so that the authors could develop their subject and provide a reliable image of their style of thought. Correlatively, the volume should not exceed the limits allowed by Springer. Beyond these material constraints, we wanted the authors to be representative both in terms of national and international recognition. We wanted the texts to be genuinely original, and provided in English. We did not want the papers to be comments on “French philosophy of science”, but rather a sample of significant contributions reflecting what philosophy of science is today in France. We also took seriously the publisher’s concern that the book be thematically coherent and representative, and truly constructed, rather than a series of free contributions. With this in mind, we elaborated a provisional outline of the volume that was sent to all prospective authors, with the names of contributors, and a relatively precise definition of the topics that we proposed to them, in function of their own work. This method excluded the possibility that several papers dealt with the same topic. But it raised a serious difficulty, because it resulted in not inviting several colleagues for whom we had high esteem. Out of the 19 authors contacted, one did not answer because of illness, two declined because they did not want to be labeled as “philosophers of science” (although major international figures in that respect), and three renounced because they were not able to provide their paper in time. We decided not to contact new authors, first because we did not want to modify the list of contributors that had been sent to the invited authors, and, secondly because the overall result fitted well with the general outline of the volume, except for philosophy of logic and mathematics, which constitutes an obvious gap of this book.

The book is divided into six parts. The first two parts deal with general issues in philosophy of science. Part I treats of styles in philosophy of science. Anne Fagot-Largeault offers a classification of methodological styles in philosophy of science at an international level. Daniel Parrochia presents a classification of styles in philosophy of technology with particular emphasis on the French case. In Part II, entitled “General philosophy of science”, Anastasios Brenner and Sandra Laugier examine two major traditional problems, theory choice and realism, and analyze the interactions, analogies and differences between important American and French philosophers on these subjects (Duhem, Meyerson, Bachelard, Quine, Kuhn).

The four other parts of the book bear upon subjects related to special scientific areas. Part III contains two essays in philosophy of physics, and one in philosophy of chemistry. Cohen-Tannoudji and Zwirn, both physicists and philosophers, characterize modern physical theory, with a different approach: Cohen-Tannoudji offers a rational reconstruction of the intellectual itinerary of 20th physics; Zwirn favors a discussion of the issue of realism in the physics of today. The third chapter

of Part III is devoted to chemistry, a scientific discipline that inspired important contributions by French or French-speaking philosophers of science in the 20th century (especially Duhem, Meyerson, Metzger, Bachelard, Stengers). Bensaude-Vincent relies on these authors, and on more recent authors worldwide, to contend that chemistry requires “a philosophy of its own”, based upon the recognition of the irreducibly technical aspect of chemistry (“knowing through making”), and the subsequent necessity for philosophers of taking history of science seriously.

Part IV deals with the philosophy of life sciences. François Dagognet, known for his important work in the history and philosophy of medicine, claims that pharmacy and pharmacology should be more seriously considered by philosophers, because of the inextricable mixture of science, technology and moral problems exhibited by its history. Claude Debru summarizes his own philosophical work on the physiological sciences (especially paradoxical sleep and the classification of leukemia), and defends “the idea of philosophy as an interpretative, reflective and speculative activity which can be practiced within science as well as about science”. Jean Gayon offers a characterization of “philosophy of biology” as it has developed in the past 30 years approximately at an international level.

Parts V and VI bear upon human sciences, or at least, sciences for which humans are central: behavioral sciences, cognitive sciences and economics. Joëlle Proust, who was until recently the only French philosopher to engage in the modern debate on function, a debate that has raged in philosophy of biology and philosophy of psychology since the early 1980s, supports a theory of function alternative to the two traditional “systemic” and “etiological” theories of function, and applies it to the notion of “mental function”. Daniel Andler offers a broad picture of the origins, scope and structure of cognitive science. Andler denies that philosophy of science should aim only at solving foundational and methodological issues, or particular conceptual puzzles: “there is also the more general concern of providing a perspective on the structure and dynamics of a field, its relations to other areas of inquiry, its purported limitations or misconceptions, its future directions”. Part VI is entirely devoted to economics, with an essay by Philippe Mongin, who convincingly shows that Duhem’s methodological holism (rather than the radical version of epistemological holism defended by Quine) provides a fascinating tool for examining the testability (or rather untestability) of expected utility theory.

How far do these contributions illuminate the question of whether contemporary French philosophy of science has a distinctive character? As we already said, the purpose of this book is not to give a direct answer to this question. We did not ask the authors to confront it, but rather to write an essay on a topic of their own, and one most able to illustrate the character of *their* style of thinking for an international audience. Most of our contributors asked us whether they should offer a review of French literature on the question. Our response was that they were free to do so if they thought it necessary, but that this was not a requirement: they were free to treat their subject with the method and with the help of the literature they felt appropriate.

Nevertheless, the present introduction is an exception to the rule of the book. The first two sections endeavor to describe the particular paths taken by French

philosophy of science in the 19th and 20th centuries, and offer a few conjectures about the possible coherence of this history. In this third section, we cannot avoid drawing some lessons from the contemporary contributions that we have received. Of course, we are aware that our sample is quite restricted, and probably biased, although we have attempted to be as open to various schools and networks as possible. But precisely because we did not ask the authors to make a statement on “French philosophy of science” and to defend it, we can the more freely venture to extract from our sample several tendencies that were not predictable from the onset, given the rules we had proposed to our authors. Readers, especially non-French readers will be free, of course, to draw their conclusions from their own perspective, and we do expect that this kind of dialogue will arise. As far as we are concerned, four general tendencies can be seen in the material that we have gathered.

First, one may ask how many authors felt the necessity of explaining what philosophy of science in general is or should be. Eight authors have explicitly faced this issue, and clearly considered it a decisive issue, intimately related to their way of practicing philosophy of science. This is not a surprise in the case of the first chapters, dealing with “Styles in Philosophy of Science and Technology”. But a similar concern, with one exception, appears in all the other parts of the book: Brenner in Part II, “General Philosophy of Science”, Bensaude-Vincent in Part III, “Physical and Chemical Sciences”, Dagognet, Debru and Gayon in Part IV, “Life Sciences”, and Andler in Part V, “Philosophy of the Behavioral and Cognitive Sciences”. All agree that history (history of science and technology, but also history of philosophy and social history) matters to philosophy of science. All of them seem also to be of the opinion that, with or without history, there is more to philosophy of science than just foundational, methodological or particular conceptual puzzles in this or that area of scientific investigation. Andler (already quoted), summarizes nicely a rather general feeling among these authors, a feeling that reminds us of Auguste Comte: one of the legitimate ends of philosophy of science is to provide “a perspective on the structure and dynamics of a field”, either from within the chosen field or by reflecting on its relations with technology, society and history by and large.

Of course, there is nothing extraordinary in such positions. Other national traditions in philosophy of science would most certainly include spontaneous assertions about the issue of what philosophy of science may or should be. Once again, this was not explicitly requested of the authors. More than half of them did it, and not only those committed to general philosophy of science. This may be related to rather deep doubts and conflicts concerning the nature and limits of philosophy of science in France today. Furthermore, all these authors insist on the importance of history – something that would probably not be expected in the case of many other countries, especially those that significantly contribute to the international literature in the field of philosophy of science.

Our second observation is about the relative importance of the human sciences. Mere counting of the number of pages devoted to the philosophy of human sciences tells us something. Of course, as editors, we tried to keep the authors of these sections (psychology, cognitive science, economics) within the assigned limits. We were surprised, however, by the density and the personal commitment of these

papers, and we finally decided not to authoritatively impose abridgement. These contributions testify to the importance of this domain in contemporary philosophy of science in general, and more especially in France. Let us note that the last three chapters of the book are totally alien to any kind of hermeneutics, historicism, relativism or social constructivism. The three authors concerned are more or less analytically inclined, and have nothing to do, for instance, with schools of thought inspired by philosophers such as Michel Foucault or Paul Ricœur, not to speak of Pierre Bourdieu, Gilles Deleuze or Jean Baudrillard (we are aware here that we point towards very different styles of thought). Nevertheless, we may observe that their work arises in a national context where the status of the human and the social sciences have always been a major concern for the French intellectual community, and especially for French philosophers of science (again, the historical example of Auguste Comte remains here as a touchstone).

Our last two observations aim at locating the sample of papers that we have collected relative to general trends in philosophy of science today worldwide. It is often said that since the mid 1960s, philosophy of science has experienced two major shifts, the “historical turn” and the “regionalist turn”. These two shifts went in the same direction: they testified to a certain skepticism regarding the idea of a general and timeless theory of science. It is worth asking here how far this volume confirms or not these tendencies in the French case. As for the historical turn, we said earlier that more than half of the authors in this volume explicitly defend the claim that “history matters to philosophy of science” in one way or another. Thus, in a sense, contemporary French philosophy of science illustrates the general tendency. But two reservations should be made. First, for most of these authors, the dictum “history matters to philosophy of science” is not the result of a “shift”. Although Thomas Kuhn has been as popular in France as everywhere else (as reflected in this volume by Brenner’s and Laugier’s papers), the interest of philosophers of science for history is rooted in an earlier tradition, which in reality was so powerful in France throughout the 20th century that Kuhn was received more as a confirmation than as a revolution – even though this idea may rest on serious misconceptions about Kuhn’s thought. Secondly, the overall picture given by the present book is definitely that of a rather deep disagreement about the role of history in philosophy of science. Although they are a minority, the papers by Zwirn, Proust, Andler, Mongin, *are not* historically inclined. Furthermore, had we taken a larger sample of philosophers of science, especially including younger scholars, we would have observed in all likelihood a strong tendency toward non-historical work, both in the fields of general and special philosophy of science. We may also note a hardening of the rivalry (or divorce?) between the two schools of thought, the historically inclined and the analytically inclined. This conflict, of course, has nothing exceptional, but it is particularly harsh today in France, because of both the strength (and fertility, we should add) of the traditional historico-critical school and the vigor of the new analytically oriented one.

As for the regionalist turn, things are clearer. Out of the 13 chapters of this book, nine belong to special philosophy of science, which therefore seem to weigh more in our country than general philosophy of science. It could be objected that this was

the result of the editors' choice. However, we think that our choice reflects honestly (though very partially) the best of philosophy of science in this country. It also echoes a long-term orientation of French philosophy of science underlined in the previous sections of this introduction. Since the French do not equate "epistemology" with the theory of knowledge, but with a critical reflection on the sciences as they historically existed or now exist, room for a purely general and normative philosophy of science has always been more restricted in France than elsewhere. This is not to say that few people are interested in general philosophy of science in this country. On the contrary, many of them are. But, with the significant exception of an increasing number of analytical philosophers, general philosophers of science tend to treat their subject with the spirit and methods of history of philosophy, a discipline which goes on occupying the most central role in academic training in philosophy in France. In a sense, the four papers in general philosophy of science included in this volume illustrate this tendency.

This being said, the common grid of the historical regionalist turn might well be a poor criterion for assessing the distinctiveness or non-distinctiveness of French contemporary philosophy of science. In the first chapter of this volume, Anne Fagot-Largeault suggests that three heterogeneous traditions have for a long time coexisted under the name of "philosophy of science": "formal philosophy", "historical epistemology", and "philosophy of nature". Formal philosophy is analytic philosophy applied to problems and methods belonging to science. Historical epistemology is based upon the principle that the genesis of problems, concepts, methods, and, possibly, the social structure of science, is key to a proper philosophical understanding of science. The expression "philosophy of nature" (or "natural philosophy") is taken by Fagot-Largeault in a particular sense, defined as "speculative philosophy, grounded in scientific knowledge and data, going beyond just what those known data allow to assert, with a view to seizing a unity or rationale in the ways nature is constituted". This involves both an attempt to synthesize the available scientific knowledge and, quite often, a more or less bold metaphysical reflection. According to Fagot-Largeault, this is the kind of philosophy of science that is most often developed by scientists, especially brilliant scientists (such as Whitehead, Waddington, or Schrödinger). But it is also the kind of "philosophy of science" found in general philosophers who explicitly try and develop a metaphysics inspired by positive science, and who most certainly would refuse the label "philosopher of science". Bergson is a nice example.

Then, if we accept this original taxonomy, how does it apply to the present book, and, more generally, to French philosophy of science? On the whole, formal philosophy of science and historical epistemology are both represented in this book, although not equally (see above the paragraph on the "historical turn"). The balance, after all, could change, like in other countries, in one or another direction. As for the genre of "philosophy of nature", it seems quite obvious that it constitutes an important horizon for a majority of authors in this volume. Most of them come to agree in the final analysis that the boundary between science and philosophy is uncertain. Some of them (e.g. Fagot-Largeault, Parrochia, Cohen-Tannoudji, Dagognet, Gayon, Debru) would most certainly concur with Andler that one of the

main tasks of philosophy of science, and not the least, is to provide comprehensive views of entire scientific areas and the possibility of an active dialogue with scientists in that perspective. Others (Brenner, Laugier, Bensaude-Vincent), come to the same conclusion on the basis of their more explicitly historical approach to philosophy of science.

This sympathy of French philosophers of science for “natural philosophy” (*sensu* Fagot Largeault) is probably shared with a lot of philosophers of science all over the world. But we believe that this is distinctively important in France because of the long lasting Comtian positivist tradition, which emphasizes so much the intrinsic value of science, that is to say the idea that the sciences – not a normative philosophy of science – pose their own norms of rationality. In such a historical context, one may expect that genuine scientists will play a major role on the theatre of philosophy of science. They won’t be conventionally identified as professional “philosophers of science”, but they will significantly contribute to the philosophical debate over science. For instance, to take just a few legendary examples, the mathematician René Thom, the physicists Louis de Broglie or Bernard d’Espagnat, the biologists Jacques Monod or François Jacob, have probably had a greater impact on the national and international scene of philosophy of science than most (and perhaps all) French philosophers of science of the second half of the 20th century. Again, this is not to say that similar figures do not play a similar role in other places. But we think that the French intellectual world is more inclined to produce such figures and to give them an important role in philosophy of science.

This of course is a free conjecture of our own which would require extensive comparative work to be tested. As already said, the purpose of this volume is not to offer a collection of meta-philosophical studies, but rather to give a concrete sample of what philosophy of science looks like in France today. We hope that the sample offered here will fulfill this expectation, and foster the development of fertile interaction in philosophy of science worldwide.

Acknowledgment The project of a volume on French philosophy of science arose at a conference given by Robert S. Cohen in Vienna. Recalling his editorship of the Boston Studies in Philosophy of Science series he had launched, he pointed out the unfortunate absence, among the numerous volumes devoted to different national traditions, of one on French studies in the field. We express our thanks to Robert Cohen for his suggestion and his encouragements to follow it up. We also voice our indebtedness to Jürgen Renn and Kostas Gavroglu, for their acceptance of our project in the series and valuable help in bringing it to completion.

Endnotes

- 1 Prior to Comte’s explicit definition of the field, there were indeed endeavors that opened the path to philosophy of science, for example the *Encyclopédie* (1751–1780) of Diderot and d’Alembert or Laplace’s *Essai philosophique sur les probabilités* (1814).
- 2 In his advertisement, dated 1829, Comte writes: “I was careful not to chose the denomination of natural philosophy nor that of philosophy of science [*philosophie des sciences*], which would have been perhaps even more precise, because neither the one nor the other yet apply

- to all orders of phenomena, whereas positive philosophy, by which I understand the study of social phenomena as well as all the others, designates a uniform way of reasoning applicable to all subject matters upon which the human mind can exert itself", 1930–1942, vol. 1, p. VI. For a lexicological study of the expression "philosophy of science" in French, English and German, see H. Pulte, 2004.
- 3 Whewell responded to Comte's endeavor, in particular in the second edition of his *Philosophy of inductive sciences*, IV, 6. We know that J.S. Mill engaged more directly with Comte in his *System of logic* published shortly thereafter in 1843.
 - 4 On Canguilhem's relation to Comte, see his 1968. Serres is one of the editors of the current edition of Comte's *Cours de philosophie positive*. Cf. Serres, 1972, 1974.
 - 5 See Comte, 1844, § 12. Cf. Neurath, 1983.
 - 6 See Comte, 1830–1842, vol. 1, p. 464. Whewell, 1840, II, p. 315.
 - 7 Letter to W. James of the 9th of May 1908, in Bergson, 1972, p. 776; 2002, p. 199. Bergson in all likelihood is thinking of the term *épistémologie*, which had gained currency and had lately been used with insistence by Meyerson.
 - 8 Milhaud went so far as to speak of "positivisme logique" as early as 1905. He thereby refers to the conception of Renouvier and perhaps his own early attempt to reformulate it. Milhaud, 1927, p. 55, reproducing articles published in 1905.
 - 9 Le Roy, speaking of determinism, writes "This belief assumes that science is *an adequate knowledge of reality*; whereas, for modern epistemology [*épistémologie moderne*], its aim is merely man's action on nature", in *Actes du Premier congrès international de philosophie, Revue de métaphysique et de morale*, 8, 1900, p. 540. In the same volume Russell takes epistemology as philosophy of knowledge, p. 562.
 - 10 Despite the fact that the term had been coined earlier in English, French philosophers insisted on referring to science, resorting to *gnoséologie* eventually to designate the theory of knowledge.
 - 11 Three years later, in 1895, a chair in philosophy of inductive sciences was established at the University of Vienna for Ernst Mach.
 - 12 See Janik and Toulmin, 1973.
 - 13 The official name of the chair becomes "History and philosophy of science".
 - 14 This institute continues to exist today. Now called "Institut d'histoire et de philosophie des sciences et des techniques", it associates the Université de Paris I, the Centre national de la recherche scientifique and the École normale supérieure. For details on its founding, see Braunstein, 2006.
 - 15 Bachelard, 1934, p. 178. Cf. Koyré, 1966, p. 14.
 - 16 Bachelard, 1934, p. 178.
 - 17 Bachelard, 1934, p. 40.
 - 18 See Bachelard, 1932 and 1933.
 - 19 Bachelard, 1928, pp. 185–191.
 - 20 Canguilhem, 1943.
 - 21 Canguilhem, 1966.
 - 22 Canguilhem, 1989.
 - 23 For an explicit argument in favor of this view of philosophy of science, see Canguilhem 1968.
 - 24 This aspect of Canguilhem has been commented on again and again by a huge number of French philosophers of science. See however the external appraisal by Grene, 2000.
 - 25 Foucault, 1985. See also Rabinow, 1994.
 - 26 Neither Bachelard or Canguilhem coined the term "historical epistemology", which was introduced by Dominique Lecourt (1969), with reference to Bachelard, in the context of a master's thesis written under Georges Canguilhem. Today, many authors tend to think that the genre of historical epistemology, that is epistemology based upon careful historical work, is better illustrated by Canguilhem. On these issues, see Gayon in Gutting, 1989, Gayon 2003, Rheinberger 2005, Lecourt 2008.
 - 27 Dagognet, 1967, 1970.

- 28 Dagognet 1964.
 29 See Brittan, 1991.
 30 Vuillemin, “Ma vie en bref” in Brittan, 1991, p. 4.
 31 See Proust and Schwartz, 1995.
 32 See for example *La parole malheureuse* (1971), *La force de la règle* (1987), *Le pays des possibles* (1988).
 33 See Comte, 1830–1842, Table of Contents.
 34 Ernest Renan, 1890, chap. 16, p. 301.
 35 Duhem, 1906, p. XV; translation, p. 3. Cf. Duhem, 1913, p. 115.
 36 See above n. 26.
 37 See Bachelard, 1951, chap. 1; 1972, “L’actualité de l’histoire des sciences”. Cf. Canguilhem, 1968, pp. 181–184.
 38 For a study in English on Foucault, see Gary Gutting, 1989.
 39 For example the activities of the International Society for the History of Philosophy of Science (HOPOS) and of the Vienna Circle Institute. See M. Heidelberger and F. Stadler, 2002.
 40 See for example Jocelyn Benoist, 1999 and Pierre Wagner, 2002.
 41 See Canguilhem, 1952; Simondon, 1969; Dagognet, 1977; Latour, 1992.
 42 See Lecourt, 1990, 2003.
 43 Duhem, 1906, p. 199; translation, p. 133, modified.
 44 On Nicod, Cavaillès and Hebrand, see Bitbol and Gayon, 2006.
 45 This is not to mention the efforts of mathematicians: Paulette Destouche-Février did a lot to introduce to France research in logic after World War Two. Her own work dealt with non-classical logic in relation to quantum mechanics. Georg Kreisel, an Austrian logician trained in Cambridge, initiated a research program in combinatory logic during the 1960s in Paris, which was taken up by Jean-Louis Krivine and became a genuine school.
 46 They were intent on providing a careful presentation of logical positivism and analytic philosophy, but did not dismiss history of science.
 47 We wish to express warm thanks to Michel Bourdeau, director of research at the Centre national de la recherche scientifique (IHPST Paris), for providing us with helpful information concerning logic in France.
 48 Brenner, 2003, 2006; Gayon, 1994, 1998a, 1998b, 2003, 2005; Bitbol and Gayon, 2006; Gayon and Wunenburger 2000.
 49 <http://www.sps.ens.fr/>

Bibliography

- Ampère André-Marie (1834–1843), *Essai sur la philosophie des sciences*, 2 vol., Paris, Bachelier.
 Bachelard Gaston (1928), *Essai sur la connaissance approchée*, Paris, Vrin, 1987.
 Bachelard Gaston (1932), *Le pluralisme cohérent de la chimie moderne*, Paris, Vrin, 1973.
 Bachelard Gaston (1933), *Les intuitions atomistiques: Essai de classification*, Paris, Vrin, 1975.
 Bachelard Gaston (1934), *Le nouvel esprit scientifique*, Paris, PUF, 1971; *The New Scientific Spirit*, trans. A. Goldhammer, Boston, Beacon Press, 1984.
 Bachelard Gaston (1951), *L’activité rationaliste de la physique contemporaine*, Paris, PUF.
 Bachelard Gaston (1972), *L’engagement rationaliste*, Paris, PUF.
 Benoist Jocelyn (1999), *L’a priori conceptuel: Bolzano, Husserl, Schlick*, Paris, Vrin.
 Bergson Henri (1972), *Mélanges*, Paris, PUF.
 Bergson Henri (2002), *Correspondances*, Paris, PUF.
 Bitbol Michel, Gayon Jean (2006) (eds), *L’épistémologie française: 1830–1970*, Paris, PUF.
 Bouveresse Jacques (1971), *La parole malheureuse, de l’alchimie linguistique à la grammaire philosophique*, Paris, 1987.

- Bouveresse Jacques (1987), *La force de la règle: Wittgenstein et l'invention de la nécessité*, Paris, Minuit.
- Bouveresse Jacques (1988), *Le pays des possibles: Wittgenstein, les mathématiques et le monde réel*, Paris, Minuit.
- Braunstein Jean-François (2006), "Abel Rey et les débuts de l'Institut d'histoire des sciences et des techniques (1932–1940)", in M. Bitbol and J. Gayon (eds), 2006, pp. 173–191.
- Brenner Anastasios (1990), *Duhem: science, réalité et apparence*, Paris, Vrin.
- Brenner Anastasios (2003), *Les origines françaises de la philosophie des sciences*, Paris, PUF.
- Brenner Anastasios (2006), "Quelle épistémologie historique? Kuhn, Feyerabend, Hacking et l'école bachelardienne", *Revue de métaphysique et de morale*, 1, pp. 113–127.
- Brittan Gordon (1991), *Causality, Method and Modality: Essays in Honor of Jules Vuillemin*, Dordrecht, Kluwer.
- Canguilhem Georges (1943), *Essai sur quelques problèmes concernant le normal et le pathologique*, Publications de la Faculté des Lettres de l'Université de Strasbourg, Clermont-Ferrand, Imprimerie "La Montagne," 2nd ed., Paris, Les Belles Lettres, 1950.
- Canguilhem Georges (1966), *Le normal et le Pathologique*, Paris, PUF; *The Normal and the Pathological*, trans. Carolyn R. Fawcett in collaboration with Robert S. Cohen with an introduction by Michel Foucault, New York, Zone Books, 1989.
- Canguilhem Georges (1968), "L'objet de l'histoire des sciences", in *Études d'histoire et de philosophie des sciences*, Paris, Vrin.
- Canguilhem Georges (1952), *La connaissance de la vie*, Paris, Vrin, 1975.
- Canguilhem Georges (1968), *Études d'histoire et de philosophie des sciences*, Paris, Vrin, 1970.
- Comte Auguste (1830–1842), *Cours de philosophie positive*, M. Serres, F. Dagognet A. Sinaceur and J.P. Enthoven (eds), new ed. revised, 2 vols, Paris, Hermann, 1998; *The Positive Philosophy of August Comte*, trans. and condensation H. Martineau, 1853.
- Comte Auguste (1844), *Discours sur l'esprit positif*, A. Petit (ed.), Paris, Vrin, 1995.
- Dagognet François (1964), *La raison et les remèdes*, Paris, PUF.
- Dagognet François (1967), *Méthodes et doctrine dans l'œuvre de Pasteur*, Paris, PUF.
- Dagognet François (1970), *Le catalogue de la vie: étude méthodologique sur la taxinomie*, PUF.
- Dagognet François (1977), *Une épistémologie de l'espace concret: néogéographie*, Paris, Vrin, 2006.
- Duhem Pierre (1906), *La Théorie physique, son objet, sa structure*, Paris, Vrin, 1981; *The Aim and Structure of Physical Theory*, trans. P.P. Wiener, Princeton, Princeton University Press, 1991.
- Duhem Pierre (1913), *Notice sur les titres et travaux scientifiques*, Bordeaux, Gounouilhou.
- Foucault Michel (1985), "La Vie: l'Expérience et la Science," *Revue de Métaphysique et de Morale*, special issue on Canguilhem, vol. 90, pp. 3–14.
- Gayon Jean (1994), *Bachelard: le rationalisme appliqué*, Paris, CNED.
- Gayon Jean (1998a), "Dagognet: quelle philosophie pour quelle histoire des sciences?", in *François Dagognet, médecin, épistémologue, philosophe. Une philosophie à l'œuvre*, R. Damien (ed.), Paris, Institut Synthélabo pour le progrès de la connaissance, pp. 59–77.
- Gayon Jean (1998b), "The Concept of Individuality in Canguilhem's Philosophy of Biology", *Journal of the History of Biology*, 31, pp. 305–325.
- Gayon Jean, Wunenburger Jean-Jacques (2000) (eds), *Bachelard dans le monde*, Paris, Presses Universitaires de France.
- Gayon Jean (2003), "Bachelard et l'histoire des sciences", in *Bachelard et l'épistémologie française*, J. J. Wunenburger (ed.), Paris, Presses Universitaires de France, pp. 51–113.
- Gayon Jean (2005), "Bergson Spiritualist's Metaphysics and the Sciences", in *Continental Philosophy of Science*, Garry Gutting (ed.), Malden/Oxford/Victoria, Blackwell, pp. 43–68.
- Grene Marjorie (2000), "The Philosophy of Science of Georges Canguilhem: A Transatlantic View", *Revue d'histoire des sciences*, vol. 53, n 1, pp. 47–64.
- Gutting Gary (1989), *Michel Foucault's Archeology of Scientific Reason*, Cambridge University Press.
- Heidelberger Michaël, Stadler Friedrich (2002) (eds), *History of Philosophy of Science*, Dordrecht, Kluwer.

- Janik Allan, Stephen Toulmin (1973) *Wittgenstein's Vienna*, New York, Touchstone.
- Koyré Alexandre (1966), *Études d'histoire de la pensée scientifique*, Paris, Gallimard, 1973.
- Latour Bruno (1992), *Aramis ou l'amour des techniques*, Paris, La découverte; *Aramis or the Love of Technology*, trans. C. Porter, Cambridge, Mass., Harvard, University Press, 1996.
- Lecourt Dominique (1969), *L'épistémologie historique de Gaston Bachelard*, Paris, Vrin, 2002.
- Lecourt Dominique (1992), *L'Amérique entre la Bible et Darwin*, Paris, PUF.
- Lecourt Dominique (1999) (ed.), *Dictionnaire d'histoire et philosophie des sciences*, Paris, PUF.
- Lecourt Dominique (2003), *Humain, posthumain: la technique et la vie*, Paris, PUF.
- Lecourt Dominique (2008), *Georges Canguilhem*, Paris, PUF.
- Milhaud Gaston (1927), *La philosophie de Charles Renouvier*, Paris, Vrin.
- Mill John Stuart (1843), *A System of Logic*, in *Collected Works*, J.M. Robson (ed.), vols 7, 8, London, Routledge, 1996.
- Neurath Otto (1983), *Philosophical Papers (1913–1946)*, trans. R.S. Cohen and M. Neurath, Dordrecht, Reidel.
- Proust Joëlle, Schwartz Elisabeth (1995) (eds), *La connaissance philosophique: Essais sur l'oeuvre de Gilles-Gaston Granger*, Paris, PUF.
- Pulte Helmut (2004), "Wissenschaftstheorie; Wissenschaftsphilosophie", in J. Ritter et al. (ed.), *Historisches Wörterbuch der Philosophie*, 12 vols, Basel, Schwabe, 1971–2005.
- Rabinow Paul (1994), "Introduction: a Vital Rationalist," in F. Delaporte (ed.), *A Vital Rationalist. Selected Writings from Georges Canguilhem*, New York, Zone Books, pp. 11–22.
- Renan Ernest (1890), *L'avenir de la science*, Paris, Calman-Lévy, 1925.
- Rheinberger Hans-Jörg (2005), "Reassessing the Historical Epistemology of Georges Canguilhem", in G. Gutting (ed.), *Continental Philosophy of Science*, London, Blackwell.
- Serres Michel (1972), *Hermès II: Interférences*, Paris, Minuit; *Hermès: Litterature, Science, Philosophy*, trans. of selected essays J.V. Harari and D.F. Bell, Baltimore, Johns Hopkins University Press, 1982.
- Serres Michel (1974), *Hermès III: La traduction*, Paris, Minuit.
- Simondon Gilbert (1969), *Du mode d'existence des objets techniques*, Paris, Aubier-Montaigne.
- Wagner Pierre (2002) (ed.), *Les philosophes et la science*, Paris, Gallimard, 2002.
- Whewell William (1840), *The Philosophy of the Inductive Sciences*, in G. Buchdahl and L.L. Laudan (eds), *The Historical and Philosophical Works of William Whewell*, 10 vols, Frank Cass, London, 1967.

Part I
Styles in Philosophy of
Science and Technology

The Legend of Philosophy's Striptease (Trends in Philosophy of Science)

Anne Fagot-Largeault

The title is meant to tease the reader and attract his/her curiosity, but the question behind the teasing is serious. The reader will gently excuse the unconventional gait of a chapter that originated as an invited lecture given in Paris, at the HOPOS 2006 June conference. Doing philosophy of science requires having been trained both in philosophy and in (at least some) science. That is already a challenge. Studying the history of philosophy of science (which is what “hopos” means) might require having been trained as a historian as well. As life is short, and no one is omniscient, philosophy of science and its history can only be the endeavour of a community of researchers. A common endeavour calls for, if not a plan, at least a common rationality. What follows is about doubts and hopes, and about the reasons we have for tolerating, and even loving, a variety of styles in the ways philosophy of science is practiced.

A rough survey of notorious works in philosophy of science will suggest (at least) three different styles. The argument goes through five points, with the subtitles: 1. Science and philosophy, 2. The legend, 3. Beyond the legend, 4. Styles in philosophy of science, 5. New questions, emerging styles?

1 Science and Philosophy

Is the relation between science and philosophy internal or external? Does science belong to philosophy, is philosophy inherent in science, are science and philosophy independent of each other? Let us here contrast two philosophers who were contemporaries, with very divergent views on the relation between science and philosophy.

A. Fagot-Largeault (✉)
Collège de France
11 place Marcelin Berthelot
75005 Paris, France
E-mail: anne.fagot-largeault@college-de-france.fr

1.1 “Science Generates Philosophy” (Bachelard¹)

The French philosopher (and physicist) Gaston Bachelard would encourage young philosophy students to read scientific publications, to get to know researchers in science, to go to the laboratory and see them work, or even to do laboratory work themselves. He called the plunging into science at work an experience of “meeting double transcendence”, that is, an experience of exposing oneself to be sanctioned both by hard facts, and by the criticism of other members of the group. He was convinced of two things. First, numerous philosophical questions emerge from science itself, and those are real questions, that is, questions about how the world really is. Second, philosophers should risk conjectures that are vulnerable to refutation or correction by scientific results, rather than taking refuge in unfalsifiable “intuitions”. Bachelard indeed thought that “*science does not get the philosophy it deserves*”², because philosophers tend to indulge in autistic speculation, while scientists too often satisfy themselves with primitive metaphysical beliefs. The philosophy science deserves would be a “discursive metaphysics open to rectification”³.

1.2 “Science Does not Think” (Heidegger)

The German philosopher Martin Heidegger is well aware of the provocative tone of his claim when, in a series of lectures that were given during the Winter semester of the academic year 1951–1952, he declares:

Science does not think, and cannot think; indeed, that is what constitutes its chance, that which secures its own way of proceeding. Science does not think. A shocking assertion. Let it be shocking, even if we complete it with another assertion: that science is always somehow related – in its peculiar way – to thinking.⁴

This does not merely assert that practicing scientists do not have time for leisurely philosophizing. What Heidegger means is that there is a gap, a fracture between science and philosophy. It is common knowledge that he ascribes the fracture to the technological drift of contemporary science: when you want results, you cannot think. (Then, only in the old times, when science meant contemplation, when it aimed at pure, disinterested knowledge, only then, perhaps philosophy and science were one and the same?) But there is more to Heidegger’s charge against the quest for power. The leitmotiv of the 1951–1952 course is: “up until now, Man has done too much, and thought too little” (*ibid*). Acting keeps mankind from thinking. What’s thinking? Meditating. A solitary journey. Note that Heidegger adds that even philosophers may publish a lot and think very little. At any rate, an activist science is foreign to philosophical thinking.

Heidegger’s way (science is technoscience, and technoscience is unrelated to serious philosophy) may be safer than Bachelard’s (scientific investigation is philo-

sophical by its nature). The poet and essayist Paul Valéry, who enjoyed the company of scientists and philosophers, lucidly warned them:

In our times metaphysics was seen to be caught by surprise in the variations of science, and eventually bullied in the most hilarious manner. That's why I'd happen to think that, if I were a philosopher, I would endeavour to make my philosophical reflection independent of any knowledge that new experience might shatter.⁵

2 The Tale of Philosophy's Striptease

In the late 20th Century many a philosopher feared that philosophy was about to disappear, at least from university education. Philosophy is in danger, they said, it needs to be rescued. Or they would tell, as did Stephen Toulmin with a touch of humour, "how medicine saved the life of ethics"⁶, and did not save the life of epistemology... Where did the threat come from?

2.1 *Philosophy and the Breeding of Science*

The story goes like this. Note that it should not be mistaken as history. It is a tale. Once upon a time philosophy included all sciences, and the technologies derived from science, and the wisdom that goes with such endeavours. A well-known representation of such a concept of philosophy is Descartes' tree of knowledge:

The whole of philosophy is like a tree, the roots of which is metaphysics, the trunk is physics, and the branches issued from the trunk are all other sciences, which come down to principally three, namely mechanics, medicine and ethics; I mean, the highest and most perfect ethics, based on full knowledge of all other sciences, thus being the ultimate degree of wisdom.⁷

The tree eventually lost its branches. Sciences fell off philosophy and became autonomous. First, the sciences of nature, in the course of the 17th and 18th centuries: physics, chemistry, biology, with Boyle, Newton, Lavoisier, Lamarck, etc. Then, the noosciences, or sciences of culture, in the course of the 19th century: psychology, and the social sciences. Philosophy was then faced with a trilemma: either rehearse its own history and lament over its being stripped of all the good sciences it carried, or claim to be one of the sciences, or take refuge in literature and poetry – which was Paul Valéry's option, when he said that a piece of philosophy isn't any more serious than a piece of music.⁸

The legend may have been borne as a misunderstanding of W. Dilthey's reflection on the emergence of human sciences as historical sciences. We will come back to Dilthey later. At this point, let us ask a question. What happens after the tree has lost its branches, that is, after philosophy has been delivered of all the sciences? What is left?

2.2 What is “Pure” Philosophy, in the Nude?

Obviously the roots of the tree are still there. Shouldn't philosophers be busy enough with *metaphysics*, considering (as Kant said, quoting a latin poet) that “nothing is over as long as something remains to be done”⁹? But whoever has read the *Critique of Pure Reason* (1781, 1787) to the end has been persuaded that metaphysical constructions are hopeless. Even though Kant calls metaphysics a “*fundamental science*”, he maintains that the scope of metaphysics has to be strictly limited, following a serious analysis of the limits of human knowledge. Metaphysical schemes are without substantive import. They may have, at the most, heuristic or speculative value, but they do not tell us anything about the real world.

Indeed, Immanuel Kant had been captivated by *prima philosophia*'s appeal. This he confessed in *Dreams of Spirit-Seer* (1766):

My fate was to fall in love with metaphysics, although I could hardly pride myself on having been granted her favors.¹⁰

Such a seductive metaphysics was like Swedenborg's chimeras, or at best Plato's world of Ideas, an enchanting world, a dream! A rational critical attitude required to chase away the dream and face the facts.

2.3 Interpreting the Tale: Edmund Husserl vs. Bertrand Russell

Quite a few philosophers in the early 20th century apparently accepted the story of philosophy being robbed and deprived of scientific disciplines that had originated from her. Husserl deplores the fact, Russell is delighted.

According to Edmund Husserl, the scientific *impulse* – the want for good knowledge, for clear, sure, valid, apodictic knowledge – lay within philosophy. The “vital task” of philosophers is to build a sound and universal science, based on rock-firm ground, totalizing all truths about the world. Unfortunately, as explained in *Philosophy as rigorous science*, while striving to be scientific, philosophy merely succeeded in giving birth to independent sciences:

The only ripe fruit of such efforts has been to establish in their independence the rigorous sciences of nature and mind, as well as the new branches of pure mathematics.¹¹

What happened was that, as they started accumulating a body of knowledge, the sciences underwent a process of *naturalization*: first the sciences of nature, then even the sciences of consciousness, became *positive*. They got engulfed in the object, and lost track of the founding subject giving birth to the data. The sciences of nature are rigorous in that they are critical of their experimental procedures, but they do not criticize their foundations. It sounds as if this setback were, to some extent, philosophy's fault. However, once naturalized, the sciences are incomplete and not truly autonomous. The objective is to bring them back into philosophy's womb. The tree of knowledge must be reassembled:

There is only one philosophy, only one true and genuine science, within which the particular genuine sciences precisely are non-autonomous members.¹²

Unlike Husserl, Bertrand Russell finds philosophy's striptease most enjoyable. He welcomes philosophy getting denuded of her belongings because he does not believe in the foundational role of philosophical reflection. Philosophy is idle chat, vacuous talk. We need the precise tools of mathematical logic to clarify philosophical problems, and eventually solve them, or else discard them as insolvable, or pointless. All philosophical claims have to be systematically reexamined, and as a result of stringent analysis, philosophy ends up as empty as a puppet:

I believe the only difference between science and philosophy is, that science is what you more or less know and philosophy is what you do not know ... Therefore every advance in knowledge robs philosophy of some problems which formerly it had, and if there is any truth, if there is any value in the kind of procedure of mathematical logic, it will follow that a number of problems which had belonged to philosophy will have ceased to belong to philosophy and will belong to science.¹³

This was written during Russell's "Wittgensteinian" epoch (as admitted by Russell himself, in *The Monist*, 1918). The reader is reminded of assertions found in the *Tractatus Logico-Philosophicus*.¹⁴

3 Beyond the Legend

Science does not belong to philosophy: let us admit it. The claim that it does is, at least to some extent, a delusion. From that it does not follow that science is foreign to philosophy. The programme of modern empirical science, as designed by Francis Bacon, talked of "natural history" (or the investigation of facts in the universe) as a platform which would "serve for a foundation to build philosophy upon"¹⁵, and his "natural philosophy", which in the *Advancement of Learning* (1605) he calls "metaphysic", states the laws ("axioms") of nature drawn from experience by the inductive method. From that perspective, natural philosophy belongs to science, or perhaps philosophy is an extension of science. Let us now ask: how did philosophy of science emerge as a discipline?

3.1 From Natural Philosophy to Philosophy of Science

Although Bacon wants the laws of nature to be derived from scientific observation and experiment (as opposed to deducing them from "brain-created" general principles), he takes for granted that there are universal laws of nature, because nature is the well-ordered creation of God; even if "we will have it that all things are as in our folly we think they should be, not as seems fittest to the Divine wisdom, or as they are found to be in fact", he warns that "we cannot command nature except by obeying her"¹⁶. Most of the early natural philosophers, including Newton,¹⁷ share this assumption.

In the course of the 19th century, however, the transition from natural theology to natural philosophy places Nature and its dark spontaneity as an alternative to God. Different regions of nature may then have their specific regularities, or Nature may vary. “Any science must have its own philosophy”, Lamarck writes,¹⁸ speaking of zoology. Doing natural philosophy in general, *i.e.* philosophy of Nature, in the manner of John Herschel,¹⁹ and relying on inductive methods, takes a bet on Nature’s consistency. Auguste Comte admits that his “positive philosophy” is much like what is called “natural philosophy” in England, but his *Course*²⁰ is segmented according to a hierarchy of natural sciences, from physics to sociology (from more general and simpler, to more particular and complex), with a view to evidence both the transitions between disciplines and the proper “scientific genius” of each science.²¹ Charles Darwin admired Paley’s *Natural Theology*,²² but there is a vast distance from the harmonious world of Paley to a Darwinian order.

The diversity of sciences calls for philosophical questioning. William Whewell in 1840 publishes his *Philosophy (singular) of the Inductive Sciences (plural)*.²³ Whewell does not have a system of the sciences, as Comte does; he considers several kinds of sciences, each with their core concepts or Ideas, including sciences (such as geology) in which the present state of things is explained by past events, a type of research he qualifies as “aetiological history”. The phrase “philosophy of the sciences” is commonplace around the middle of the 19th century. It is replaced by “philosophy of science” at the end of the 19th century, at the expense of philosophy becoming what Pearson²⁴ terms “the grammar of science”. Although Pearson takes his examples from physics, on account of the unity of its “grammar” the unity of science is postulated for all “branches of knowledge”. To sum up: the 19th century takes us from the unity of nature to the unity of science, and from philosophy of nature to philosophy of science. Meanwhile there is considerable diversification of the sciences, due to the more and more irreducible diversity of objects studied.

3.2 *Antoine-Augustin Cournot and the Non-Scientific Status of History*

Antoine Augustin Cournot agrees with the distinction (coming from F. Bacon, slightly modified by d’Alembert in the *Encyclopaedia*²⁵) between two types of knowledge: historical, and theoretical.²⁶ Indeed, some aspects of historical research may be considered scientific, but other aspects are irreducibly unscientific. The scientific part of the work consists in establishing the facts and their chronology. The narrative part, namely the reconstruction of the course of events, involves speculative choices (drawing up causal links, distinguishing between chance events and rational links within clusters of events); the uncertainty of such choices cannot be eliminated, the probability of their being right is “philosophical probability”, and “philosophy of history” resides in the speculation they involve.

Scientific knowledge proper is theoretical: formulate general hypotheses or theories, and submit them to the “criterion of experiment”, or demonstration, is its

motto. Scientific theorizing, however, may go beyond what can decisively be settled by way of experiment. The speculative attempt at going beyond the facts already known and looking for hidden regularities, or at “capturing analogies and searching for the reasons of things”, is risky, for it amounts to “wandering from those facts which can be rigorously controlled”²⁷. Cournot uses the terms “philosophy of the sciences” to mean: “philosophical speculation inherent in scientific work”. He reckons that there is in mathematics, physics, biology, economics,

a part of philosophical speculation, from which science cannot refrain, or, should it cut off from it – assuming that were possible – it would be at the expense of its own dignity.²⁸

In either case (be it history, or science) the speculative part of the research is deemed philosophical, although it is a philosophy from inside science, and possibly made by scientists. Cournot himself, a mathematician by training, and the discoverer of a grand theorem in mathematical economics (the Cournot–Nash equilibrium), considered philosophical speculation a natural continuation of scientific investigation.

3.3 *W. Dilthey and the “Essence” of Philosophy*

Wilhelm Dilthey ambioned to establish the noosciences, or sciences of the mind and culture (*Geisteswissenschaften*) as an autonomous group of sciences, distinct from the sciences of nature (*Naturwissenschaften*). He was aware of German philosophers, from Schelling to Hegel and Schopenhauer, trying persistently (against Kant) to provide a metaphysical foundation for the sciences:

So the possibilities of the metaphysical method were tried in Germany, one after the other, and always with the same negative result.²⁹

Dilthey rejects both Kantian transcendentalism and postkantian foundationalism. He does not like J.S. Mill's naturalistic solution either. In the *System of Logic*, Mill claims that the “backward state” of what he names the “moral sciences” should be remedied by applying to them the methodology of physical science. Dilthey wants the specificity of moral science to be preserved. In his *Introduction to the Human Sciences*,³⁰ he argues that individualizing those sciences does not require to rely on such old metaphysical distinctions as that of material *vs.* spiritual substances. It is enough to refer to the distinction between our experience of the external world, and our inner experience, and characterize for the two types of experience the appropriate scientific approach: explain phenomena in terms of cause and effect relations in the former case, possibly explain *and* especially understand in the latter case.

In fact, Dilthey analyzes at great length³¹ the ways in which modern sciences, starting with physical sciences, had to get emancipated from *prima philosophia*, or the metaphysics of substantial forms. That may be one of the sources of the legend of philosophy being deprived of the sciences. If so, the legend was based on a misinterpretation. Dilthey does not mean to say that modern sciences are in no relation to philosophy, he means to say that they are *not rooted* in philosophy. Human sciences

have their roots in human (real) history. They do not need metaphysical principles for a start, they need a methodology.

Philosophy is not the foundation of human science: it is its achievement, its “ultimate result”. Science culminates in philosophy:

The set of mind sciences combined all historical research under the viewpoint of universal history, rooted those sciences in history, and gathered together philology, criticism, historiography, comparative history and history of evolution. That way history turned into philosophy.³²

4 Styles in Philosophy of Science

Thomas Kuhn was interested in analysing the “essential tension”³³ within scientific research between “divergent thinking” and “convergent thinking”. This section will deal with the tension within philosophy of science induced by various (historically attested) ways of viewing the relation between philosophy of science. Schematically, divergent interpretations given to the “tale” yield several ways of conceiving the tasks of a philosophy of science.

Assume science has severed its links with philosophy (science “does not think”). “Pure” philosophy stands by herself. It may ignore science. It also may, from outside, take science as an object for study, among other objects. As there is a sociology and a psychology of science, there will be a grammar of science (if philosophy’s expertise is grammar).

Now assume that science has rejected metaphysical preconceptions and kept its philosophical momentum (science “thinks”). There remains a special link between philosophy and science, a secret complicity, no matter how critical the partners get of each other. What do they share? It may be a common desire for truth. How do they differ? One may hypothesize that science works at conquering new pieces of knowledge, and philosophy retrospectively studies how science did the work (philosophy accompanies science, as history and methodology of science); then philosophy and science complement one another. One may also hypothesize that theoretical speculation, prospective thinking, from inside science, is tentatively philosophical.

Let us briefly examine the three styles just sketched.

4.1 *Formal Philosophy of Science*

Famous textbooks illustrate what Susan Haack calls the “linguistic-conceptual-analytical style” which was dominant within the English-speaking philosophical community during the 20th century. Arthur Pap’s *Introduction to the Philosophy of Science* is a superb example; it also has the merit of warning that there exist other styles.

The philosophy of science is here conceived as indistinguishable from analytic philosophy except that the analysis is restricted to concepts and problems that are especially rel-

evant to science. It should be distinguished from a different conception of the philosophy of science as a speculative synthesis of the fruits of scientific research. ... Logic courses overlap a great deal with philosophy of science-in-general.... Epistemology can hardly be distinguished from philosophy of science-in-general, provided its problems are problems of logical justification of beliefs and not (psychological) problems of genetic explanation of beliefs³⁴

Philosophy here stands by itself as an “analytic” discipline, the tools of which are essentially logic and grammar. Science is identified with its linguistic productions, namely scientific publications. When a philosopher is interested in a piece of science, his/her task is to detect the primary notions, check that other notions are properly defined in terms of the primary ones, look for the axioms of the theory (or axiomatize the theory), make sure that the set of axioms is consistent and eventually that the axioms are independent of each other, interpret the axioms and build models, verify that the author’s conclusions are properly derived from the axioms, or pinpoint the flaws in the argument.... In short, philosophical expertise is with language, the tool for language analysis is logic, philosophy identifies with logic. Logical treatment has to do with the structure of knowledge, not with its contents (supposedly the form does not add anything substantial to the contents). In that context, philosophy is not a source of knowledge. The philosopher is a sort of constable, a public officer in charge of restoring or maintaining law and order. Genuine philosophical problems are problems of logic.

Bertrand Russell anticipated such a conception of philosophy, when he bluntly declared that logic is “the essence of philosophy”:

Every philosophical problem, when it is subjected to the necessary analysis and purification, is found either to be not really philosophical at all, or else to be, in the sense in which we are using the word, logical.³⁵

In other words, philosophy is either nothing, or a science; and philosophy of science must be science of science. Rudolf Carnap, Hans Reichenbach,³⁶ symbolize the migration of “scientific philosophy” from Berlin and Vienna to the United States before World War II.

An inquiry into the relationship between formal methods and philosophical investigations was carried out in 2005 by editors of the *Synthese* journal.³⁷ A bunch of living philosophers, who had been trained in that style of philosophy, were asked why they were initially drawn to formal methods, and what role formal methods had played in their philosophical work. The outcome helps to understand what was achieved along those lines. Most philosophers trained in logic and mathematics did get results; for example, Adolf Grünbaum praises the “axiomatic rigor” by which he could demonstrate that a number of Euclid’s “theorems” did not follow from their premises, unless additional postulates were made explicit. There are, however, reservations about an exclusive use of formal methods in philosophy of science. Sven Ove Hansson mentions the “dangers of oversimplification”, Timothy Williamson resented the “abuse of formalization” in the manner in which Davidson’s programme was taught at Oxford, Jaakko Hintikka warns us that there is inseparability of form and substance and that it makes no sense to develop a formalism without worrying about its substantive purpose, Susan Haack learned formalism and turned to pragmatism:

Pragmatism opened my eyes to a conception of philosophy broader and more flexible than, as Tony Quinton [1983] puts it, the “lexicographical needlework” of pure linguistic analysis.³⁸

The concerns expressed by formal philosophers of science are foundational and critical, rather than constructive. They mean to dissect, clarify, justify, make sure that science proceeds on firm ground. Clark N. Glymour observes that philosophers are oversensitive to incoherence and tend to obstinately go back to first premises, while scientists tend to go ahead and be insensitive to contradiction. Moreover, formal analysis is (like mathematical truth) timeless and non-historical. Patrick Suppes had already pointed out “a tension between those who advocate historical methods as the primary approach in the philosophy of science and those who advocate formal methods”³⁹. Suppes himself is very tolerant of such a diversity of approaches: “This tension in itself is a good thing. It generates both a proper spirit of criticism and a proper sense of perspective”⁴⁰. Now what is, according to Suppes, the specific contribution of formal philosophy of science?

For me ... the best way to think about formal methods in philosophy is in providing a foundation for mathematics or for the sciences. I include in the sciences the problem of clarifying the foundations of probability.⁴¹

Ruth Barcan Marcus thinks of logicians interested in other sciences as essentially contributing their critical mind:

Philosophers who are not ignorant of work in other disciplines ... have also proved to be incisive critics. There has been and could be fruitful collaboration as in linguistics, law and cognitive science.⁴²

Finally, Gabriel Sandu remembers that the Finnish philosopher Georg Henrik von Wright, a formal philosopher himself, anticipated a change of style at the turn of the 21st century:

In this very department [Helsinki], shortly before his death, von Wright predicted that this century will not be that of logic, as the preceding one, but that of speculative philosophy.⁴³

As a matter of fact, formal methods are alive and well, not only in English speaking countries, but also on the European continent, possibly with a preference for applications to the general theory of knowledge, rather than for the study of particular sciences. An example is a recent doctoral dissertation on “Propositional attitudes and epistemic paradoxes”⁴⁴.

4.2 History and Philosophy of Science, and Historical Epistemology

The historical epistemology (or HPS) approach is rooted in a strong tradition on the European continent, from Comte and Cournot to Duhem, Koyré and Canguilhem. It occasionally migrated to the other side of the ocean. Its concerns are with the genealogy of scientific discoveries, the building of scientific concepts and theo-

ries, the methodologies of establishing scientific facts and/or justifying scientific assertions, the modes of scientific explanation, etc. It counts many famous works, of which we may mention just a few: Ernst Mach's *Science of Mechanics* (1883),⁴⁵ Leon Brunschvicg's *Stages of Mathematical Philosophy* (1912),⁴⁶ Emile Meyerson's *Explanation in Science* (1921),⁴⁷ Pierre Duhem's majestic *History of Cosmological Doctrines* (1913-1959),⁴⁸ Ludwik Fleck's *Genesis and Development of a Scientific Fact* (1935),⁴⁹ Gaston Bachelard's *Formation of the Scientific Mind* (1938),⁵⁰ Georges Canguilhem's *Formation of the Concept of Reflex* (1955),⁵¹ Alexandre Koyré's *Astronomical Revolution* (1961),⁵² Ernst Mayr's *Growth of Biological Thought* (1982),⁵³ Lorenz Krüger's (et al.) *Probabilistic Revolution* (1987),⁵⁴ etc. All of these works combine (internal) history of science and a philosophical investigation into what genuine scientific knowledge and research is and should be. For example, Mach warns that his critical-and-historical approach is "antimetaphysical"; Meyerson finds (chemical) science to be "essentially ontological", that is, he dismisses radical positivism; Mayr argues for life science requiring an epistemological paradigm different from that of physics.

Thomas Kuhn, who was drawn "from physics and philosophy to history", insists that the history and the philosophy of science are and should remain "separate and distinct disciplines"⁵⁵. Georges Canguilhem quotes Dijksterhuis saying that the history of science is "not only the memory of science, but also its epistemological laboratory"⁵⁶. In fact, Canguilhem says, "epistemology has always been historical", even in Kant's second preface to the *Critique of Pure Reason*.⁵⁷ There can't be proper epistemological study without historical contextualisation. Conversely, the choice of interesting historical matter may rely on epistemological judgement:

There are in fact two versions of the history of science: the history of obsolete knowledge and the history of sanctioned knowledge, by which I mean knowledge that played an active role in the making of subsequent science. Without epistemology it is impossible to distinguish between the two. Gaston Bachelard was the first to make this distinction.⁵⁸

The distance between "history of obsolete knowledge" and "history of sanctioned knowledge" is illustrated by two doctoral dissertations, both recently defended. One is a careful study of the medical conceptions of migraine in the 19th century;⁵⁹ it describes a succession of intellectual fashions, culturally fascinating, scientifically eccentric, and the patients went on living with their headaches. The other analyses how the concept of neuron emerged during the 20th century,⁶⁰ from a research on the nervous system previously directed at the neural fibre; it shows that a multiplicity of new technologies and new experimental strategies coming from various disciplines converged to constitute neurophysiology as the science of neural cells and neural networks; this study gives a clear sense of a gain in knowledge in the midst of multiple erring ways.

Often debated within the historical-epistemological tradition is the question, whether there is continuity or discontinuity in the evolution of science. Ernst Mach tells how the laws of mechanics were progressively discovered through human (biological) experience of motion; he suggests that there is continuity from experience to science, advocates description as demonstration, and claims that historical knowledge is necessary for proper understanding of current science. He comments:

That which for Husserl is a humiliation for scientific thinking, namely its continuity with common (“blind”?) thinking, that I see as an element of grandeur, for that is how science is rooted in the depths of human life and has a deep impact on it.⁶¹

Bachelard, on the contrary, as soon as 1938,⁶² popularized the notions of epistemological *obstacle* and epistemological *rupture*, to emphasize the discontinuity between common and scientific ways of thinking.

Cournot (around 1870) initiated the technical use, in history of science, of the concept of *scientific revolution*.⁶³ Cournot is not a radical discontinuist, although he admits of deep reorganizations of knowledge, such as the introduction of quantitative chemistry by Lavoisier, which deserve the qualification of “revolutionary”. Scientific revolutions reemerged in the 20th century, together with scientific paradigms. The notion of *paradigm* was launched by Kuhn, with reference to Fleck’s notion of *Denkstil*. It was soon followed by Foucault’s notion of *épistémè*.⁶⁴ The kind of history of science using such concepts may be viewed as a discontinuist variety of historical epistemology, with philosophical assumptions taken from structuralism (Gestalt theory) – the paradigm being a conceptual structure, incommensurability being discontinuity between structures. The group of scholars who gathered in Bielefeld during the academic year 1982-83 to study the rise of probability in the sciences did raise the question: was the rise of probability revolutionary? Bernard Cohen gives four criteria to decide when a scientific revolution has occurred. Lorenz Krüger opts for a “slow rise of probabilism”. Ian Hacking concludes that, even though one cannot talk about a probabilistic revolution in the scholarly sense, yet “the taming of chance and the erosion of determinism constitute one of the most revolutionary changes in the history of the human mind”.⁶⁵

Hacking defines himself as practicing another variety of historical epistemology, which borrows both from the analytic and the continental styles. The guide line is to combine the “Lockean imperative to investigate the origin of ideas” with Michel Foucault’s “history of the present”. As a consequence, historical enquiries

are not done out of curiosity about the past. They are intended to show something about our present reality, our present reasoning, our present modes of research.⁶⁶

Together with historian Lorraine Daston, in the Fall of 1993, Hacking had a workshop in Toronto on the topic of “historical epistemology”. The working paper written in preparation of the event enumerates what Hacking’s brand of historical (meta?)-epistemology is not: “(a) not studies of theory change ... (b) not social studies of science ... (c) not cognition, not biological ... (d) not overarching pictures of civilization ... (e) not deconstruction ... (f) not archaeology, not genealogy ... (g) not science in action”⁶⁷. This quotation helps to point out the large variety of substyles flourishing during the late 20th century, within what is commonly designated as HPS. It may be fun for the reader to try and guess who lies behind each designation.⁶⁸

4.3 *Philosophy of Nature/Philosophical Anthropology*

At the end of his life, Christiaan Huygens wrote a philosophical treatise, entitled *Cosmotheôros*, in which, on the basis of recent astronomical discoveries, he con-

jectures how the universe may really be. Far from being spherical and limited, it is immensely vast. It contains a great many other suns than our sun, with other planets moving round them, and those planets might be habitable:

So that what we allowed the planets, upon the account of our enjoying it, we must likewise grant to all those planets that surround that prodigious number of suns. They must have their plants and animals, nay and their rational creatures too, and those as great admirers, and as diligent observers of the heavens as ourselves⁶⁹

Huygens' *Cosmotherôs* is an early piece of modern natural philosophy, just like Freud's *Civilization and its Discontents*⁷⁰ is an early contemporary piece of philosophical anthropology, a conjectural essay on the uncertain outcome of the conflict (evidenced by psychoanalytic research and the social sciences) between human native aggressivity and its cultural domestication.

Natural philosophy is speculative philosophy, grounded in scientific knowledge and data, going beyond just what those known data allow to assert, with a view to seizing a unity or rationale in the ways nature is constituted. Such a speculation remains, at least for a part, vulnerable to refutation or correction by the findings of further scientific investigations. The enterprise is risky, to the extent that it is both tentatively metaphysical, and fallible. The French philosopher Jean Largeault finely analysed the metaphysical patterns between which natural philosophy tends to oscillate: form vs. matter, mechanism vs. dynamism, determinism vs. contingency, continuity vs. individuation, etc.

I wanted to approach the enigma of forms, which let themselves be perceived in the physical world, which mathematicians discover through direct or abstract intuition, while their union with matter, or in Lautman's words their incarnation, remains incomprehensible.⁷¹

Some philosophers, like Bergson, have ambioned to develop a *positive metaphysics*, the intuitions of which would be empirically controllable. Bergson, however, denies to metaphysics the role of an "hypothetical extension of science"⁷²; he wants for his metaphysics a separate access to experience. In fact, even though Bergson's book on *Creative Evolution* (1907) might qualify as *philosophy of nature*, most philosophy of nature is the work of scientists trying to synthesize by building bridges between several scientific domains, or to anticipate the directions of further scientific research. August Weismann held the chair of zoology at the university of Freiburg. Towards the end of the 19th century he speculated that unicellular organisms are potentially immortal, and that the division of labor between somatic and germ cells in multicellular organisms (a product of natural selection) ends up in the mortality of the soma; he postulated a continuity of the germ plasm, which implied that acquired characters could not be inherited; his *Essays upon Heredity*⁷³ are very "philosophical" in character, yet many of his intuitions were eventually confirmed by cytological investigations.

Natural philosophy flourished in the early 19th century. Engulfed under the wave of positivism, it then became marginal, until it reappeared in the 20th century as an urge for philosophy from inside science. Erwin Schrödinger, in the Preface of his *What Is Life?*, confesses having inherited a "longing for unified, embracing knowledge", an excuse for the physicist to imagine a molecular picture of the "hereditary code-script" by which living beings reproduce themselves. When Conrad H. Waddington, in 1960, was invited to give a series of lectures in biology at the University

College of the West Indies, in Kingston, his first lecture was on “the natural philosophy of life”, in which he explains that biological research has both a “craftsman aspect”, implying an “effort to control the world”, and a speculative aspect, expressing a “wish to understand the world”, which precisely he calls *natural philosophy*. Waddington was an early promoter of systems biology. He pays homage to the mathematician and physicist Alfred N. Whitehead for having “dehorned” the main dilemma of theoretical biologists: should biology concentrate on the constituents or on the architecture of living beings? In other words, is the organism reducible to the sum of its parts, or is the investigation of architecture another avenue for discovering more about the constituents? Waddington holds that, in a sense, “the architecture is more important than the constituents”⁷⁴. Whitehead himself reckons “that the world is unfathomable in its complexity, and that anything you put together ... ought to be open to criticism if it is any good at all”. He calls his philosophical endeavour an *Essay in Cosmology*, “cosmology” is his word for “philosophy of nature”, and he defines it as speculative philosophy:

Speculative philosophy is the endeavour to frame a coherent, logical, necessary system of general ideas in terms of which every element of our experience can be interpreted.⁷⁵

Scientific philosophy, or philosophy of science? Jacques Monod in his *Chance and Necessity*⁷⁶ takes it for granted that there is a philosophy inherent in modern biology (philosophy within science), which merely has to be made explicit; for example: “all living beings are at the same time fossils”, carrying within the most minute part of their structure the marks of their common ancestry. The mathematician René Thom assumes that dynamic living phenomena, such as the development of embryos, or in general morphogenetic processes, obey mathematical constraints, which may be modeled using the (qualitative) tools of topology; the philosophical import of his quest for such “archetypes”, according to Thom, is to offer a strictly monist conception of living beings, “dissolving the mind and body antinomy into a unique geometrical entity”⁷⁷. Trained in physics and astronomy, Eric Chaisson suggests a general scheme of universal evolution, apt to synthesize astrophysics and bio-chemistry, and to “reconcile the theoretical destructiveness of thermodynamics with the observed constructiveness of cosmic evolution”. His ambition is much like that of Whitehead:

Cosmic evolution is a search for principles that subsume, and even transcend, Darwinian selection – a unifying law, an underlying pattern, or an ungoing process perhaps, that creates, orders, and maintains all structure in the Universe, in short a search for a principle of cosmic selection.⁷⁸

Of philosophical anthropology, as it developed in Europe approximately between 1920 and 1960, one may also ask whether it is a philosophical way of doing science, or a scientific style of philosophizing. The central idea, shared by von Uexküll, Buytendijk, von Weizsäcker and others, is that in doing research on living beings, including humans, one is dealing with other subjects, and not merely with objects for study. Von Weizsäcker in 1939 offers a conjecture that is intended to guide the research in such sciences; he calls it *Gestaltkreis* (cycle of structure). It is a general scheme of the unity of receptivity and movement: perception calls for action and

action calls for perception, the passive and the active revert to each other. In the second edition of his book *The Gestaltkreis*, von Weizsäcker notes that what he meant to say has since been brilliantly expressed by a French philosopher named Sartre (with his duality *en soi/pour soi*). Sartre had probably read or heard of the German book (he does not mention it). Von Weizsäcker was both a physician trained in neurology and internal medicine, and a researcher with a laboratory besides his clinic at the University of Heidelberg. His clinical practice was innovative, he introduced psychosomatic techniques, he had a theory of the patient both enduring (passively) and managing (actively) her disease. Finally, he developed a general philosophy of what it is to be human, that he calls a theory of the *split person*: “a human being is a thing (to be) *linked with* a subject (to feel, to be moved)”⁷⁹. That is undoubtedly speculative philosophy based on neuroscience and clinical experience.

At a meeting of the French Philosophical Association, in 1928, there was a controversy between Gabriel Marcel and Léon Brunschvicg. The disputed question was: Why not write novels, rather than speculating on the basis of scientific data? Marcel's opinion was that the rich fantasy of literary imagination is far more exciting than speculation restricted to what would be compatible with science. Brunschvicg answered:

L.B. What is richer in concrete reality: the universe of imagination or the universe of scientific intelligence? Descartes and Spinoza raised the question. Their answer is straightforward: the universe of imagination is fragmented and its gaps allow for miracles, whereas mysteries fade away as the study of cosmic movements within the double vastness of space and time reveals the unity and solidarity of the real world.⁸⁰

To be sure, writing a doctoral dissertation in the natural philosophy style requires a strong scientific background, and philosophical boldness, but some candidates have recently met the challenge. An experienced physicist defended a thesis in which he proposed a readjustment of Whitehead's cosmologic scheme on the basis of new developments in quantum mechanics.⁸¹ A medical doctor and pediatric surgeon developed a philosophy of pain, strongly rooted in, and very critical of, current (pseudo?) science and practice implying an ontological dualism of psychic and physical pain.⁸² In the latter case, the philosophical questioning evidently came from inside clinical science and practice.

5 New Questions, Emerging Styles?

While standard styles are still operational, at least in France, as evidenced by the topics chosen by doctoral students for their dissertations, quite a few samples of mixed styles have surfaced, and new questions are springing up from science itself. Patrick Suppes declared in 1979:

The tyranny of any single approach or any single method, whether formal or historical, should be vanquished by a democracy of methods that will coalesce and separate in a continually changing pattern as old problems fade away and new ones arise.⁸³

That is pretty much what we observe today: a plurality of methods, mixed styles and some new problems.

5.1 “We Need a Philosopher”

A survey of doctoral dissertation topics in France shows that philosophy of science dissertations are a small minority (5%) among philosophy PhDs: the overall majority is history of philosophy. But in the HPS area philosophy has at times been *called for*. Finding that there is a social demand for philosophy of science is gratifying.

Every other year the United Nations publish a *World Population Prospective*. It has become common knowledge that all countries in the world will sooner or later have to face the problem of their population’s aging. The French government in 2007 asked a group of experts to devise an “Alzheimer plan”. Experts were researchers in neuroscience, and in social science, clinicians, nurses and social workers, economists, hospital managers. They went around asking: “we want a philosopher”. A philosopher expert in Alzheimer’s disease? There was one. A doctoral student was in the process of defending his dissertation on “Philosophical problems raised by Alzheimer’s disease (history of science, epistemology, ethics)”⁸⁴. He was most welcome and helpful in the group. Why had he chosen such an exotic research topic? He happened to know a geriatrist and a biologist working in the field, both had expressed perplexities: how could a variety of senile dementia have virtually become the whole of dementia, what sense did it make to investigate animal models of human dementia, how to respect the autonomy of a person who, as dementia progresses, is losing her autonomy? Those were philosophical questions, issued from a bio-medical milieu, taken up and worked out by a philosopher.

New problems have indeed arisen in and about science. Firstly, scientific research is more and more technological, nay industrial, and scientists have developed an awareness (eventually a fear) of their *creative power*. Secondly, the growing and crucial dependence of nations on science and technology for access to basic commodities (water, health, energy) goes with a social control of scientific research (*via* funding or legislation), and a questioning (nay distrust) of scientific expertise, which has deeply modified what used to be called *freedom of research*, and has led to the development of an expertise in the evaluation of research projects, prior to the research being conducted. Thirdly, scientific *rationality* is more and more collective in character, a puzzle for philosophers.

5.2 Science is Creative

Philosophers of science tend to take for granted that science produces statements about the world *as it is*: hypotheses, theories, scholarly papers published in professional journals. It may be the case that “true” statements in physics and astrophysics aim at describing the real world as we observe it. But as soon as 1860, when Berthelot launched “synthetic chemistry”, it became clear that chemistry does more than that. Chemistry, said Berthelot, is not only about analysing the natural properties of simple or compound bodies made by nature (as Lavoisier said), it is also about

synthesizing new compounds that nature did not actualize: "Chemistry creates its object of study. Such a creative power is analogous to the power of art; it essentially distinguishes chemistry from natural & historical sciences"⁸⁵. Thus chemical science investigates not only *existing* substances, but also *possible* substances. Chemists do not produce statements only, they enrich the world with new *beings*. Their creativity interacts with nature's spontaneous creativity.

There is ample evidence to show that biology for half a century now has followed the same path, revealing possibilities that were implicit in nature. The phrase "synthetic biology"⁸⁶ was introduced in 1980 to mean the use of recombinant DNA technology to produce genetically engineered bacteria. More recently it refers to efforts at *redesigning life*, that is, isolating interchangeable modules in living systems and reassembling them in non-natural ways, with a view to clarify aspects of their functioning which are not easily accessible by mere analysis. Claude Debru parallels the "tinkering around" of human biotechnology and that of natural evolution: researchers take advantage of possibilities that are inherent in nature, they exploit and combine them, they go by trial and error, as does natural evolution; eventually also they freeze at the perspective that they might have triggered off a disaster. Debru's confidence in natural trends has a Leibnizian flavor:

Common judgement considers some technological innovations as extraordinary. It has not yet been accustomed to perceive their conditions of possibility, inherent in the mechanisms and phenomena of biological evolution. Possibilities spontaneously tend to actualize because they are partially actualized already.⁸⁷

In brief: modern biology creates new beings, new entities which did not previously exist, and which come to inhabit our world, such as genetically modified rice or maize, transgenic tomatoes, chimerical mice. The acceptability of creating, for the purpose of research, hybrid embryos (*e.g.* human sperm, mouse oocyte) or cybrid embryos (transfer of a human nucleus into an enucleated animal oocyte), was hotly debated in England in 2007. The possibility of deriving gametes (sperm and oocytes) from embryonic stem cells, and from those gametes to build "artificial embryos" successfully developing (at least) up to the blastocyst stage, has been established for mice; a similar breakthrough is likely to be achieved sooner or later for human gametes. Our ways of making babies, our conception of parenthood, are being shattered. The problems inherent in such advances cannot escape philosophers of life science. Might they claim that those are mere ethical or political problems, while philosophy of science is essentially theoretical? There is no golden standard keeping philosophy of science from being practical, or testifying to a higher dignity of theoretical philosophy over practical philosophy. Can it be argued that practical problems are the business of philosophers of technology, while philosophers of science should inquire into knowledge processes? Gilbert Hottois has convincingly argued that the acquisition of knowledge nowadays is highly technological, that there is no "pure" scientific knowledge (even in mathematics) completely isolated from techno-science, and that there is in our technoscience a "technopoiesis", or even a "technopoetry"⁸⁸, well worth the philosopher's attention.

If scientific research is indeed creative, philosophers of science have a right, perhaps a duty, to examine the possibilities opened by science, discriminate between

these possibilities those that they would or would not wish to see actualized, justify their choices. Especially when it comes to human genetics and anthropotechnology, there is space for history of discoveries, speculation on their potential impact on our image of humanity, careful study of the intertwining of methodological and ethical requirements in research strategies, etc. Recent doctoral dissertations in France illustrate a renewal of interests and a diversity of mixed styles. There was an essay on the technological remodeling of the human body (exploration of actual possibilities, criteria for judgement): “From biomedicine to anthropogeny. Epistemological and ethical reflection”⁸⁹; a questioning of the notion of *genetic fate*, in the context of prenatal or preimplantation diagnosis of hereditary diseases: “Free choice and individual fate. Concepts and problems within predictive medicine”⁹⁰; an incisive analysis of the impact of risky transplantation techniques: “What is at stake in liver transplantations with living donors”⁹¹.

5.3 *Freedom of Research vs. Human Dignity*

It is disquieting to observe that in Europe it has been, so far, impossible to get a consensus on which type of research should, or should not, be permitted in life sciences. Indeed, there have been examples of a world consensus on the prohibition of research on weapons of mass destruction (*e.g.* biological), of which it is well known that it was secretly disobeyed. The point here lies elsewhere. In France the use of the technique of somatic cell nuclear transfer has been prohibited by law, and the researchers who would use it would be heavily punished. Ireland and Italy are on the same line. In the United Kingdom the Human Fertilization and Embryology Authority has given explicit permission to some research groups to use that technique. Sweden and Belgium would be inclined to go the British way. What is at stake? Human dignity. Do we have different conceptions of human dignity, or a different appreciation of the technique? Such a question is of interest to philosophers of science.

The *Universal Declaration on the Human Genome and Human Rights* (UNESCO, 1997) states that reproductive cloning of human beings is “contrary to human dignity” (Art. 11), and should not be permitted. It does not say that the technique of nuclear transfer is objectionable in itself. It also says that “States should take appropriate steps to foster the intellectual and material conditions favourable to freedom in the conduct of research ...” (Art. 14). On the other hand, it is specified that “in the case of research, protocols shall ... be submitted for prior review” (Art. 5). Authorities in charge of scrutinizing the research protocols will therefore have the responsibility of deciding whether the research is compatible with “respect for the human rights, fundamental freedom and human dignity ...” (Art. 10).

A doctoral student had the opportunity to observe such an authority at work. It is a committee in charge of examining research protocols in epidemiology, the advice of which conditioned the granting of a permission to use, for the purpose of the research, nominative lists of sensitive data about people’s health. The commit-

tee's basic assumption was that using the data to serve a bad protocol would be an insult to the dignity of the subjects. As a consequence, the scientific quality of the protocols had to be evaluated. And the members of the committee had developed an expertise at discriminating reliable protocols from poor ones:

How to identify a phony research protocol which complies only apparently with scientific requirements? What are the exact criteria allowing to distinguish good science from trash science? Is there a consensus on what a bias would be, which compromises the credibility of a research protocol? Such questions do not arise when one writes the history of mature science, of scientific endeavours having survived a multiplicity of selection procedures.⁹²

The doctoral dissertation is about demarcation between science and pseudo-science in that context. It identifies three kinds of demarcation criteria (epistemic, non-epistemic, peri-epistemic), and shows how the *ethos* of epidemiological research defines itself. What is new here is that the (complex) criteria are extracted from an empirical observation of epidemiologists at work, instead of being conceived a priori.

5.4 *The Collective Character of Scientific Rationality*

Ludwik Fleck pointed to the "style of thought" inside science, and to the dependence of individual scientists on a "thought collective"⁹³. This was interpreted by Kuhn as prefiguring the notion of "paradigm". Perhaps there is more to it than that. No individual researcher today can assimilate the whole of science, even within her own domain of research. Within large research projects, each individual contributes her particular expertise, and *trusts* her partners for their expertise. At the individual level, scientific knowledge is in fact a mixture of knowledge, belief and trust. The overall rationality of an enterprise such as the Human Genome Project is dependent on the collective functioning of a group. Detailed studies are needed, at the crossroads between sociology of organizations, logic and psychology of human interactions, and collective ethics, of the ways various expertises adjust. In our "philosophy of science" book we had a chapter on "the intersubjective construction of scientific objectivity"⁹⁴. In his essay on the evolution of reason Bertrand Saint-Sernin⁹⁵ analyses the social conditions, the epistemological basis and the anthropological foundations of the new rationalism.

The scientific community is fragmented: each field scrutinizes one sector, from one perspective, with its own methods; but the scientific community has learned how to work collectively, and pays attention to junctions between domains, making it possible to eventually reassemble pieces of knowledge with a view to maintain a coherent representation of the natural world. The community of philosophers of science is also fragmented; the coherence of its rationality may depend on its capacity to bring individuals to complement each other within collective research projects. Alasdair McIntyre once wrote:

The building of a representation of nature is, in the modern world, a task analogous to the building of a cathedral in the medieval world or the founding and construction of a city in

the ancient world, tasks which might also turn out to be interminable. To be objective, then, is to understand oneself as part of a community and one's work as part of a project and part of a history.⁹⁶

Endnotes

- 1 Gaston Bachelard, 1934, p. 3: "La science crée en effet de la philosophie".
- 2 Bachelard, 1953, Intr. "Phénoménologie et matérialité", § VIII: "la science n'a pas la philosophie qu'elle mérite".
- 3 Bachelard, *Ibid.*
- 4 Martin Heidegger, 1954, I, 1.
- 5 Paul Valéry, 1929, "Léonard et les philosophes": "Notre époque a vu la métaphysique surprise par les variations de la science de la manière la plus brusque, et parfois la plus comique. C'est pourquoi il m'est arrivé de penser que si j'étais philosophe, je m'attacherais à rendre ma pensée philosophique indépendante de toutes connaissances qu'une expérience nouvelle peut ruiner".
- 6 Stephen Toulmin, 1982, pp. 736–750.
- 7 René Descartes, 1644–1647, Preface: "Ainsi toute la philosophie est comme un arbre, dont les racines sont la métaphysique, le tronc est la physique, et les branches qui sortent de ce tronc sont toutes les autres sciences, qui se réduisent à trois principales, à savoir la médecine, la mécanique et la morale; j'entends la plus haute et la plus parfaite morale, qui, présupposant une entière connaissance des autres sciences, est le dernier degré de la sagesse". p. 14.
- 8 Valéry, 1929.
- 9 Immanuel Kant, 1781, Preface of the 2nd edition, p. BXXIV, 1787.
- 10 Kant, 1766, II, 2.
- 11 Edmund Husserl, 1910.
- 12 Husserl, 1929, § 103.
- 13 Bertrand Russell, 1918, § 8.
- 14 Ludwig Wittgenstein, 1921, § 6.53.
- 15 Francis Bacon, 1620, Preface 'The Great Instauration'.
- 16 Bacon, 1620, I, Aphorism CXXIX.
- 17 Isaac Newton, 1687. See Book III, General Scholium.
- 18 Jean-Baptiste Lamarck, 1809, I, chap. 2.
- 19 John F.W. Herschel, 1830.
- 20 Auguste Comte, 1830–1842. See especially Lessons 1 (1830), 40 (1836), 45 (1837).
- 21 Comte, 1830–1842, end of 45th Lesson.
- 22 Thomas Paley, 1802.
- 23 William Whewell, 1840; revised edition 1847.
- 24 Karl Pearson, 1892.
- 25 Diderot & d'Alembert, 1751, vol. 1, Preface.
- 26 Antoine-Augustin Cournot, 1851, chap. XX, § 301.
- 27 Cournot, 1851, chap. XXI, § 331.
- 28 Cournot, 1851, chap. XXI, § 329.
- 29 Wilhelm Dilthey, 1907, p. 47: "So wurden nun in Deutschland die Möglichkeiten der metaphysischen Methode erprobt - eine nach anderen, und stets mit demselben negativen Erfolg", in *Gesammelte Schriften*, vol. V, p. 47
- 30 Dilthey, 1883, in *Gesammelte Schriften*, vol. I.
- 31 Dilthey, 1883, vol. II.
- 32 Dilthey, 1910, in *Gesammelte Schriften*, Bd VII.

- 33 Thomas S. Kuhn, 1977, II, 9.
- 34 Arthur Pap, 1962.
- 35 Bertrand Russell, 1914, "Logic as the Essence of Philosophy", Lecture 2.
- 36 Hans Reichenbach, *The Rise of Scientific Philosophy*, 1951.
- 37 Vincent F. Hendricks & John Symons, 2005.
- 38 Hendricks & Symons, 2005, 9.
- 39 Patrick Suppes, 1979, pp. 16–27. Note that Thomas Kuhn's book, *The Essential Tension*, had been published in 1977.
- 40 *ibid.*
- 41 Hendricks & Symons, 2005, p. 20.
- 42 Hendricks & Symons, 2005, p. 14.
- 43 Hendricks & Symons, 2005, p. 17.
- 44 Paul Egré, 2004.
- 45 Ernst Mach, *Die Mechanik in ihrer Entwicklung historisch-kritisch dargestellt*, Leipzig, 1883.
- 46 Brunschvicg Léon, *Les étapes de la philosophie mathématique*, Paris, 1912.
- 47 Meyerson Emile, *De l'explication dans les sciences*, Paris, 1921.
- 48 Duhem Pierre, *Le Système du Monde. Histoire des doctrines cosmologiques de Platon à Copernic*, Paris, 10 Vols, 1913–1959.
- 49 Fleck Ludwik, *Entstehung und Entwicklung einer Wissenschaftlichen Tatsache. Einführung in die Lehre vom Denkstil und Denkkollektiv*, Basel, 1935.
- 50 Bachelard Gaston, *La formation de l'esprit scientifique. Contribution à une psychanalyse de la connaissance objective*, Paris, 1938.
- 51 Canguilhem Georges, *La formation du concept de réflexe*, Paris, 1955.
- 52 Koyré Alexandre, *La révolution astronomique: Copernic, Kepler, Borelli*, Paris, 1961.
- 53 Mayr Ernst, *The Growth of Biological Thought. Diversity, Evolution and Inheritance*.
- 54 Krüger Lorenz, et al. eds., *The Probabilistic Revolution*, Vol. 1: *Ideas in History*, Vol. 2: *Ideas in the Sciences*, The MIT Press, 1987.
- 55 Kuhn, "The Relations between the History and the Philosophy of Science", in 1977, I, 1.
- 56 Georges Canguilhem, 1977, Intro. "Le rôle de l'épistémologie dans l'historiographie scientifique contemporaine", pp. 12–13.
- 57 *ibid.*
- 58 Canguilhem, 1968, "L'objet de l'histoire des sciences".
- 59 Esther Lardreau-Cotelle, 2007.
- 60 Jean-Gaël Barbara, 2007.
- 61 Ernst Mach, 1883, chap. IV, 4, § 11.
- 62 Bachelard, 1938, chaps 1, 12.
- 63 Cournot, 1872. For a commentary on Cournot's concept of scientific revolution, see Bertrand Saint-Sernin, 1998, chap. 7, pp. 157–166.
- 64 Michel Foucault, 1966, Preface.
- 65 Krüger et al. 1987, vol. I, § 2, 3, 4.
- 66 Ian Hacking, 1989–90.
- 67 Hacking, 1993.
- 68 (a) studies of theory change [Kuhn], (b) social studies of science [British social constructivists: Barnes, Bloor, Pickering], (c) cognition [Nozick], biological [Scot Atran], (d) overarching pictures of civilization [Crombie], (e) deconstruction [Derrida], (f) archaeology, genealogy [Foucault], (g) science in action [Bruno Latour].
- 69 Christiaan Huygens, 1698.
- 70 Sigmund Freud, 1929.
- 71 Jean Largeault, 1988, p. 17.
- 72 Henri Bergson, 'Métaphysique et science', in 1934, II, pp. 33–46.
- 73 August Weismann, 1889, chaps 2, 4.
- 74 Conrad Hal Waddington, 1961, chap. 1.
- 75 Alfred North Whitehead, 1929, corrected ed. 1978, Part 1, chap. 1, beginning of section 1.

- The previous quotation is from a speech on 'Process and Reality' given by Whitehead at Harvard on the occasion of his 70th birthday: repr. in 1947 and 1968, II, 114–119.
- 76 Jacques Monod, 1970, chap. 9.
- 77 René Thom, 1972, chap. 13, 7c.
- 78 Eric J. Chaisson, 2001, Prologue.
- 79 Viktor von Weizsäcker, 1956, "Theorie des Menschen", chap. 57 p. 391 in *Gesammelte Schriften*, vol. 10. See also 1939, in *Gesammelte Schriften*, vol. 4.
- 80 *Bulletin de la Société française de philosophie*, 1928, 28, 3.
- 81 Marc de Lacoste Lareymondie, 2004.
- 82 Henri-Keyvan Kotobi, 2007.
- 83 Suppes, 1979, concl.
- 84 Fabrice Gzil, 2007.
- 85 Marcelin Berthelot, 1876: "La chimie crée son objet. Cette faculté créatrice, semblable à celle de l'art lui-même, la distingue essentiellement des sciences naturelles et historiques", p. 275.
- 86 S.A. Benner & A.M. Sismour, 2005, "Synthetic Biology".
- 87 Claude Debru, 2003, chap. 3, p. 257: "La pensée commune juge comme inouïes certaines innovations techniques. Elle n'a pas encore appris à en percevoir les conditions réelles de possibilité dans les mécanismes et phénomènes de l'évolution biologique. Si le possible tend de lui-même à l'existence, c'est qu'il est, en partie, déjà réalisé".
- 88 Gilbert Hottois, 2004.
- 89 Jérôme Goffette, 1996.
- 90 Catherine Dekeuwer, 2006.
- 91 Valérie Gateau, 2006.
- 92 Nicolas Lechopier, 2007, p. 7.
- 93 Fleck, 1935.
- 94 Daniel Andler, Anne Fagot-Largeault, Bertrand Saint-Sernin, 2002, chap. 2.
- 95 Saint-Sernin, 2007, Part 3.
- 96 McIntyre Alasdair, 1978 "Objectivity in Morality and Objectivity in Science", vol. 3, p. 37.

Bibliography

- Andler Daniel, Fagot-Largeault Anne, Saint-Sernin Bertrand (2002), *Philosophie des sciences*, 2 vols, Paris, Gallimard.
- Bachelard Gaston (1934), *Le Nouvel esprit scientifique*, Paris, PUF, 1971; *The New Scientific Spirit*, trans. A. Goldhammer, Boston, Beacon Press, 1984.
- Bachelard Gaston (1938), *La Formation de l'esprit scientifique*, Paris, Vrin.
- Bachelard Gaston (1953), *Le matérialisme rationnel*, Paris, PUF.
- Bacon Francis (1620), *Novum Organum*, London. *Bacon's Novum Organum*, trans. P. Urbach & J. Gibson, La Salle, Open Court, 1994.
- Barbara Jean-Gaël (2007), *La constitution d'un objet biologique au 20e siècle. Enquête épistémologique et historique des modes d'objectivation du neurone*, Doctoral dissertation, University of Paris-7.
- Benner S.A. & A.M. Sismour (2005), "Synthetic Biology", *Nature Reviews/Genetics*, 6, pp. 533–543.
- Bergson Henri (1934), *La pensée et le mouvant*, in *Oeuvres*, Paris, PUF, 1970.
- Berthelot Marcelin (1876), *La synthèse chimique*, Paris, G. Baillière.
- Brunschvicg Léon (1912), *Les étapes de la philosophie mathématique*, Paris, Alcan.
- Canguilhem Georges (1955), *La formation du concept de réflexe*, Paris, PUF.
- Canguilhem Georges (1968), "L'objet de l'histoire des sciences", *Etudes d'histoire et de philosophie des sciences*, Paris, Vrin.

- Canguilhem Georges (1977), "Le rôle de l'épistémologie dans l'historiographie scientifique contemporaine", *Ideologie et rationalité dans l'histoire des sciences de la vie*, 2nd ed. revised., Paris, Vrin, 1988.
- Chaisson Eric J. (2001), *Cosmic Evolution. The Rise of Complexity in Nature*, Harvard University Press.
- Comte Auguste (1830–1842), *Cours de philosophie positive*, 2 vols, Paris, Hermann, 1975.
- Cournot Antoine-Augustin (1851), *Essai sur les fondements de nos connaissances et sur les caractères de la critique philosophique*, Paris, in *Oeuvres complètes*, vol. 2, Paris, Vrin, 1975; *An Essay on the Foundations of Our Knowledge*, trans. M.H. Moore, New York, Liberal Arts Press, 1956.
- Cournot Antoine-Augustin (1872), *Considérations sur la marche des idées et des événements dans les temps modernes*, in *Oeuvres complètes*, vol. 4, Paris, Vrin, 1973.
- Debru Claude (2003), *Le possible et les biotechnologies*, Paris, PUF, 2003.
- Dekeuwer Catherine (2006), *Liberté de choix et destins individuels. Examen des concepts et des problèmes éthiques impliqués par la médecine prédictive*, Doctoral dissertation, University of Paris 1.
- Descartes René (1644–1647), *Principia philosophiae; Les Principes de la philosophie*, in Adam and Tannery (eds), *Œuvres*, 11 vols, Paris, Vrin, 1996; *The Philosophical Writings of Descartes*, trans. J. Cottingham, R. Stoothoff, D. Murdoch and A. Kenny, 3 vols, Cambridge, Cambridge University Press, 1984–1991.
- Diderot Denis et d'Alembert Jean Le Rond (1751–1780) (eds), *Encyclopédie ou dictionnaire raisonné des sciences, des arts et des métiers*, 21 vols, Stuttgart, Frommann, 1966–1967.
- Dilthey Wilhelm (1883), *Einleitung in die Geisteswissenschaften. Versuch einer Grundlegung für das Studium der Gesellschaft und der Geschichte*, in *Gesammelte Schriften*, Teubner (online, via Gallica); *Introduction to the Human Sciences*, in *Selected Works*, R.A. Makkreel & F. Rodi (eds), vol. 1, Princeton, Princeton University Press, 1989.
- Dilthey Wilhelm (1907), *Das Wesen der Philosophie*, in *Gesammelte Schriften*, Göttingen, Vandenhoeck & Ruprecht, vol. 5; *The Essence of Philosophy*, trans. S.A. & W.T. Emery, Chapel Hill, University of North Carolina, 1954.
- Dilthey Wilhelm (1910), *Der Aufbau der geistlichen Welt in den Geisteswissenschaften*, in *Gesammelte Schriften*, Göttingen, Vandenhoeck & Ruprecht, vol. 7; *The Formation of the Historical World in the Human Sciences*, in *Selected Works*, R.A. Makkreel & F. Rodi (eds), vol. 3, Princeton, Princeton University Press, 1989.
- Duhem Pierre (1913–1959), *Le Système du Monde. Histoire des doctrines cosmologiques de Platon à Copernic*, 10 vols, Paris, Hermann.
- Egré Paul (2004), *Attitudes propositionnelles et paradoxes épistémiques*, Doctoral dissertation, University of Paris-1.
- Fleck Ludwik (1935), *Entstehung und Entwicklung einer Wissenschaftlichen Tatsache. Einführung in die Lehre vom Denkstil und Denkkollektiv*, Basel; *Genesis and Development of a Scientific Fact*, trans. F. Bradley & T.J. Trenn, Chicago, University of Chicago Press, 1979.
- Foucault Michel (1966), *Les mots et les choses. Une archéologie des sciences humaines*, Paris, Gallimard; *The Order of Things*, trans. A. Sheridan, New York, Random House, 1970.
- Freud Sigmund (1929), *Das Unbehagen in der Kultur*; Engl. tr., *The Standard Edition of the Complete Psychological Works of Sigmund Freud*, J. Strachey et al., 24 vols., London, Hogarth Press, 1953–1974.
- Gateau Valérie (2006), *Enjeux éthiques des transplantations hépatiques avec donneurs vivants (THDV)*, Doctoral dissertation, University of Paris-1.
- Goffette Jérôme (1996), *From biomedicine to anthropogeny. Epistemological and ethical reflection*, Doctoral dissertation, University of Bourgogne.
- Gzil Fabrice (2007), *Problèmes philosophiques soulevés par la maladie d'Alzheimer (histoire des sciences, épistémologie, éthique)*, Doctoral dissertation, University of Paris-1.
- Hacking Ian (1989-90), Two Kinds of 'New Historicism' for Philosophers, *New Literary History*, 21, pp. 343–364.

- Hacking Ian (1993), "Historical Epistemology, an Informal Discussion Paper by Ian Hacking", dittoed, Toronto workshop.
- Heidegger Martin (1954), *Was heisst denken?* Tübingen, M. Niemeyer; *What is Called Thinking*, trans. F.D. Wieck & J.G. Gray, New York, Harper & Row, 1968.
- Hendricks Vincent F. & Symons John (2005), *Formal Philosophy: Aim, Scope, Direction*, Automatic Press, 2005, www.formalphilosophy.com.
- Herschel John F.W. (1830), *A Preliminary Discourse on the Study of Natural Philosophy*, London, Longman.
- Hottis Gilbert (2004), *Philosophie des sciences, philosophie des techniques*, Paris, Odile Jacob.
- Husserl Edmund (1910), "Philosophie als strenge Wissenschaft", *Logos*, 1910, 1, pp. 289–314.
- Husserl Edmund (1929) *Formale und transzendente Logik*, Halle, Niemeyer.
- Huygens Christiaan (1698), *Cosmotheôros, sive De terris caelestibus, earumque ornatu, conjecturae Cosmotheôros*; English trans. *online*.
- Kant Immanuel (1766), *Träume eines Geistersehers, erläutert durch Träume der Metaphysik*, Königsberg; *Dreams of a Spirit-Seeer*; in *Theoretical Philosophy 1755-1770*, trans. D. Walford, Cambridge, Cambridge University Press, 1992.
- Kant Immanuel (1781), *Kritik der reinen Vernunft*, Riga, J. F. Hartknoch; 2nd ed., 1787; *Critique of Pure Reason*, trans. N. Kemp Smith, London, Macmillan, 1933.
- Kotobi Henri-Keyvan (2007), *Approche philosophique de la douleur dite 'physique' (ou la dualité du corps et de l'esprit mise à mal ... par la douleur)*, Doctoral dissertation, University of Paris-1.
- Koyré Alexandre (1961), *La révolution astronomique: Copernic, Kepler, Borelli*, Paris, Armand Colin.
- Krüger Lorenz, et al. (1987) (eds.), *The Probabilistic Revolution*, vol. 1: *Ideas in History*, vol. 2: *Ideas in the Sciences*, Cambridge, Mass., The MIT Press, 1987.
- Kuhn Thomas S. (1977), *The Essential Tension. Selected Studies in Scientific Tradition and Change*, Chicago, The University of Chicago Press, 1977.
- Lacoste Lareymondie Marc de (2004), *Philosophie de la non-séparabilité quantique et cosmologie de Alfred North Whitehead*, Doctoral dissertation, University of Paris-1.
- Lamarck Jean-Baptiste (1809), *Philosophie zoologique*, 2 vols, Paris, Dentu
- Lardreau-Cotelle Esther (2007), *Représentations de la migraine dans la France du dix-neuvième siècle (Concepts médicaux, outils thérapeutiques, images du corps)*, Doctoral dissertation, University of Paris-1.
- Largeault Jean (1988), *Principes classiques d'interprétation de la nature*, Paris, Vrin, 1988.
- Lechopier Nicolas (2007), *Ethique dans la recherche et démarcation. La scientificité de l'épidémiologie à l'épreuve des normes de confidentialité*, Doctoral dissertation, University of Paris-1.
- Mach Ernst (1883), *Die Mechanik in ihrer Entwicklung historisch-kritisch dargestellt*, Leipzig; *The Science of Mechanics*, trans. T. McCormack, La Salle, Open Court, 1974.
- Mayr Ernst (1982), *The Growth of Biological Thought. Diversity, Evolution and Inheritance*, Cambridge, Mass., Harvard University Press.
- McIntyre Alasdair (1978), "Objectivity in Morality and Objectivity in Science", in H.T. Engelhardt & D. Callahan (eds), *The Foundations of Ethics and its Relationship to Science*, vol. 3, The Hastings Center.
- Meyerson Emile (1921), *De l'explication dans les sciences*, Paris, Payot.
- Monod Jacques (1970), *Le hasard et la nécessité. Essai sur la philosophie naturelle de la biologie moderne*, Paris, Seuil.
- Newton Isaac (1687), *Philosophiae Naturalis Principia Mathematica*, London, J. Streater; *Mathematical Principles of Natural Philosophy*, trans. A. Motte and F. Cajori, 2 vols, Berkeley, University of California Press, 1962.
- Paley Thomas (1802), *Natural Theology: or, Evidences of the Existence and Attributes of the Deity collected from the Appearances of Nature*, London, Faulder.
- Pap Arthur (1962), *An Introduction to the Philosophy of Science*, New York, The Free Press of Glencoe.

- Pearson Karl (1892), *The Grammar of Science*, London, Scott, 1892.
- Reichenbach Hans (1951), *The Rise of Scientific Philosophy*, Berkeley, University of California Press.
- Russell Bertrand (1914), "Logic as the Essence of Philosophy", in *Our Knowledge of the External World*, London, Routledge, 1993.
- Russell Bertrand (1918), "The Philosophy of Logical Atomism"; repr. in *Logic and Knowledge*, London, G. Allen & Unwin, 1956.
- Saint-Sernin Bertrand (1998), *Cournot. Le réalisme*, Paris, Vrin.
- Saint-Sernin Bertrand (2007), *Le rationalisme qui vient*, Paris, Gallimard.
- Suppes Patrick (1979), "The Role of Formal Methods in the Philosophy of Science", in: P.D. Asquith & H.E. Kyburg, Jr. (eds), *Current Research in the Philosophy of Science*, East Lansing, MI, PSA.
- Thom René (1972), *Stabilité structurelle et morphogénèse*, Benjamin, Reading, MA.
- Toulmin Stephen (1982), "How Medicine Saved the Life of Ethics", *Perspectives in Biology and Medicine*, 25, pp. 736–750.
- Valéry Paul (1929), "Léonard et les philosophes", in *Variété, Théories poétique et esthétique, Oeuvres*, t. 1, Paris, Gallimard, 1957.
- Waddington Conrad Hal (1961), *The Nature of Life*, London, Allen & Unwin.
- Weismann August (1889), *Essays upon Heredity and Kindred Biological Problems*; 2nd ed., 2 vols, Oxford, Clarendon Press, 1891–92.
- Weizsäcker Viktor von (1939), *Der Gestaltkreis: Theorie der Einheit von Wahrnehmen und Bewegen*, Leipzig, 2nd ed. 1943, 4th ed. 1948, in *Gesammelte Schriften*, vol. 4., Frankfurt, Suhrkamp, 1986–2005.
- Weizsäcker Viktor von (1956), *Pathosophie*, Göttingen, in *Gesammelte Schriften*, vol. 10.
- Whewell William (1840), *The Philosophy of the Inductive Sciences, Founded upon their History*, revised ed. 1847, in G. Buchdahl and L.L. Laudan (eds), *The Historical and Philosophical Works of William Whewell*, 10 vols, London, Frank Cass, 1967.
- Whitehead Alfred North (1929), *Process and Reality. An Essai in Cosmology*, New York & Cambridge (separate editions); revised edition by D.R. Griffin & D.W. Sherburne, New York, Free Press, 1978.
- Whitehead Alfred North (1947), "Process and Reality", in *Essays in Science and Philosophy*, New York, The Philosophical Library; repr. Greenwood Press, 1968.
- Wittgenstein Ludwig (1921), "Logisch-philosophische Abhandlung", *Annalen der Naturphilosophie*, 14, p. 185–262; *Tractatus Logico-Philosophicus*, trans. D.F. Pears & B.F. McGuinness, London, Routledge & Kegan Paul, 1922.

French Philosophy of Technology

Daniel Parrochia

1 A Short History

Is there, actually, a French philosophy of technology? If the answer may be “yes”, we can point out that such a philosophy does not have a clear beginning. Of course we must go back to René Descartes (1596–1650) to find the first formulation of such a thought and the earliest idea that technology and applied sciences are essentially tools to understand and master the world. This matrix of all the later optimistic visions of science appears in the most famous work of the father of rationalism, *Le discours de la méthode* (part VI), when Descartes explains that, thanks to technology, man may become like master and owner of all of nature.¹ In the same way, we read in *Les principes de la philosophie*, another anthology piece saying that science is a tree, the roots of which are metaphysics, the trunk is physics, and the different branches are the three applied sciences known at the time (mechanics, medicine and morals).² These ideas are very significant in the context of the period, when the increase of population, the expansion of cities and towns, and a confident faith in the new Galilean physics made things looked rather hopeful. Friend of Villebressieu,³ a prolific engineer who made numerous inventions, spent his fortune building them and finally ruined himself, Descartes was an expert in technology, especially lifting devices, about which he wrote a notice enclosed with a letter to Huyghens:⁴ “The explanation of machines with the help of which a small force can move heavy weights”.⁵ But his knowledge of basic machines (pulley, screw, wheel or wheel, lever, inclined plane) does not make of Descartes a real philosopher of technology. As Blaise Pascal (1623–1662), who had one of the first calculators made by some craftsman and wrote a few things about it,⁶ Descartes was, above all, an actor of the technical revolution. But probably too close to it, he could not think a lot about the break it caused in culture. As a matter of fact, we only find, in

D. Parrochia (✉)
Faculté de philosophie
Université Jean Moulin Lyon 3
18 rue Chevreul
69007 Lyon, France
E-mail: daniel.parrochia@univ-lyon3.fr

Les principes de la philosophie (IV, 204), the famous metaphors comparing nature with an immense engine, which produces trees and fruits as a clock gives the time. Which does not mean that the “factory of earth and sky” may be reduced to a human trick. As Descartes wrote in the same passage “all rules of mechanics belong to physics, so everything which is artificial is also natural”. Here we must be careful and not misinterpret these words. In spite of the well-known Animal-machine thesis, we cannot read this sentence in the reverse order. The philosopher never said that nature is entirely mechanical, because the natural factory is not a human fact. The boss is God. His technology is subtler than ours, His intentions and goals are beyond our belief. So, some kind of deep-rooted finality remains a vivid ground of cartesianism, which prevents it turning into a flat materialism.

We cannot say exactly the same for Diderot and d’Alembert, happy project managers of the great *Encyclopédie*, one of the lighthouses of the 18th century, which contributed spreading the new technology in society at large. By adding to the numerous volumes of their beautiful book a supplement especially intended for craftsmen, they put within the reach of everyone the possibility of finding tools’ designs, technical instruments’ patterns and industrial drawings of new machines. So it is no exaggeration to say that Diderot, long before Mac Luhan and communication theory, may be viewed as a true theorist of the new media.⁷

Throughout the Enlightenment, a philosophy of technology began to develop as scientists progressed with their research. Vaucanson’s automata, for instance, leads Condorcet (1799) to some cursory reflections about what he called the “genius of mechanics” that is, for him, “an abstract art of composition”, the theory of which lies in Leibnizian *Analysis situs*, a science that made only little progress during the century. But the distance between technology and science was so obvious that, some years later, Poncelet still could see for himself that a big gap separated the teaching of mechanics everywhere in the schools from the various manners of applying it, even the more simple and usual ones.⁸

That it is quite difficult to find an actual existence of a philosophy of technology in the Classical Age in France may be easily understood: there could be, actually, nothing like it at the time. Strictly speaking, the term “philosophy of technology” is believed to find its origin in the work of Ernst Kapp, a 19th century German philosopher (1808–1896) who authored a book entitled *Philosophie der Technik*. Influenced by Georg W. F. Hegel and Karl Ritter, Kapp, in the late 1840s, fell out of favor with the German authorities and was forced to leave his homeland. He then emigrated to the German pioneer settlement of central Texas where he lived for two decades.⁹ Perhaps his life on the American frontier motivated his writing about technology, which was essentially for him – as for Descartes, indeed – a means to overcome dependence on undomesticated nature.

But that book could not play any role in the early decades of the 19th century, when Auguste Comte was attending the École polytechnique. Comte, not only as an engineer, but also as secretary to Saint-Simon, was very aware of recent developments in science and new technology. He knew as well the needs of modern societies, based on big industrial sectors, dense urban fabrics and powerful networks of communication. At that time, the Saint-Simonian school was so active

that Infantin, Lesseps, and some others, began to build a lot of roads, railways, canals (Suez), which were going to change the face of the world. So, Comte was well placed to note the birth of a new social class in western societies, the class of engineers, “whose special destiny was to organize the relations between theory and practice”.¹⁰ Yet, though he might recognize that philosophical observations on such an organization would have been of great interest and importance, he said he could not go into that. For a work which would embrace scientific knowledge and the corpus of the new class doctrine would be at present (the middle of 19th century) totally premature: because these intermediate doctrines between pure theory and direct practice still did not get off the ground.

For lack of such doctrines, French philosophy of technology began by some observations closely related to sciences of power, machines and principles of mechanisms, which appeared in the works of Carnot (1824) or Coriolis (1829), following those of Joule, in England, or Franz Reuleaux, in Germany. Those scientists, at that time, tried to compute the effects of mechanization, they measured the productive action of industry on matter and laid the foundations for a general theory of machines. In this context, the old idea of Bacon – to command obedience to Nature by beginning to obey it (*De dignitate et augmentis* II, 2) –, was defended again by Taine in the “introduction” to the *Essais de critique et d’histoire*, when he wrote that the most fruitful research is the one that enables the hand of man to interfere in the great mechanism of Nature, and so, clearly using it to his benefit, may introduce a big change in its working.

Some time later, Henry Le Chatelier, a disciple of Taine, brought that thesis to a successful conclusion by showing that the history of industrial development in the late 19th century never did anything else. Mining engineer by trade, and Professor of industrial Chemistry and Metallurgy, Le Chatelier wrote, at the end of his life, a book about science and industry,¹¹ in which, for the first time, was studied the development of applied sciences and the beginning of Taylorism. This modest and competent man, famous for the discovery of the principle bearing his name (the equilibrium principle of Le Chatelier¹²), consistently tried to integrate theory with practice and directed his most successful research toward the problems of industry. By his reflection on the development of industrial chemistry, he is perhaps one of the foremost philosophers of technology in France.

As opposed to this ambient widespread positivist thought, a philosopher like Bergson was to react and supported another thesis. While thinking that technology fulfills basic human needs and that man, a tool-making animal, is less an *Homo sapiens* than an *Homo faber* (so that technology must be considered as belonging to the very nature of mankind), Bergson did not support the purely materialistic understanding of technological artifacts.¹³ On the contrary, his philosophy pointed out the continuous nature of the world, this being for him the result of a spiritual energy expansion which explodes like a bomb and then falls again like a flare or a rocket in a big fireworks display. The successive drops in temperature make energy crystallize in life, matter and finally understanding, which has a lot of affinity with matter, but less with spirit and time (which is, for Bergson, “duration”). So, for this philosopher, technology plays only a minor role among the numerous and sometimes unknown possibilities of human mind.

After Bergson, we may observe that French philosophy of technology came to fork into different trends:

First, a descriptive philosophy of technology, connected with history, and made by former engineers or technicians, go on with Couffignal (1952), Laffitte (1932), Daumas (1953), de Latil (1953), Russo (1969 et 1986), Moles (1970) and then comes to Jacomy, Guillerme, Maitte or Ramunni.

Secondly, a more sociological and anthropological approach develops with Friedmann (1946), Leroi-Gourhan (1943), Schuhl (1948), and Lévi-Strauss (1962, 1985) who takes interest, as an ethnologist, in manual labor and “do it yourself” activities, Salomon (1984, 1986, 1992), Sfez, Musso, Gras, or even Breton.

Thirdly, an epistemological current makes its appearance with the Bachelardian school : Canguilhem (1967), Dagognet (1973, 1989, 1995), Simondon (1969), and Beaune (1980a, b), Séris, Debray, Chazal and Parrochia.

Fourthly, a properly philosophical sometimes ethical, and most often metaphysical mode of thinking arrives with French disciples of Heidegger (Dominique Janicaud, Bernard Stiegler) and of Michel Serres (Pierre Lévy) or a philosopher like Jean-Yves Goffi. We may also attach to this fourth current the works of Dominique Bourg, a student of Frank Tinland, or of the Belgian philosopher Gilbert Hottois, even if the writings of the latter offer a stubborn resistance to Heidegger’s treatment of the question of technology.

2 Technique and Technology: A Problem of Words

Until now, I used the expression “philosophy of technology” as if I did not know French language has two words – “technique” and “technologie” – for translating “technology”. In fact, they do not have the same meaning.

“Technique”, from Greek *technikos*, “which concern some art”, comes from *technè*, art, experience, skill or even ability to do something. In the former sense, it denotes a set of manufacturing processes properly applied to manual labor, especially to the use of tools as material objects, adding to the properties of human body those of natural ones (club, axe, hammer, saw, bow, sling, javelin, string, pot, net, etc.). In the modern sense, “technique” is a synonym of applied sciences, viewed as a consequence of the progress of physical sciences and of the systematic use of natural or processed forces (coal and steam, hydrocarbons, electricity, nuclear power, etc.).

“Technologie”, from Greek *technologia*, denotes at first a treatise or a description of the rules of some art. In its technical sense, the term, which is not completely fixed, is close to the second sense of “technique” and means at first technical trades (study of tools, equipment, materials, processes with a view to industrial outputs). But more and more, “technologie” is well and truly applied mathematics or physics (electronics and robotics, computer science, astronautics, satellites technology, etc.), and the complex objects produced with the help of those sciences. There is also a second sense of the word, with which “technologie” means a philosophical reflection on techniques, a study of their relations with

theoretical or fundamental sciences, the political, economical, social or moral consequences of their development, especially from the point of view of social sciences. Very often in France, philosophy of technology amounts to that. This explains that Jean-Pierre S eris¹⁴ may have said that “la technologie”, that is, a more or less general discourse on the consequences of technology, should not mask “les techniques”, that is, the art itself.

3 Technology and Science

In France, a long tradition of “history of technology” precedes (or, at least, accompanies) what is properly called “philosophy of technology”. Now, taking a certain historical view reveals many things and especially teaches us that in most cases – until the beginning of the 19th century – inventors never based their discoveries on an antecedent intellectual knowledge.

In the General Foreword of his masterly *Histoire des techniques*, Maurice Dumas clearly asserts that, for more than 20 centuries, the relationships between Science and Technology remain very fragmentary ones.¹⁵ They began certainly with the elementary contributions of astronomy and arithmetic in Antiquity, but the major scientific activity of Pericles’ century did not bring any substantial gain in the field of technology. Later, in the Middle Ages, cathedrals builders apparently borrowed nothing from mathematicians, in a time when navigation and medicine scarcely began to use scientific discoveries. It was only in the middle of the 17th century that Huyghens was able to apply isochronism pendulum oscillations found by Galileo, to the control of clocks. But this was still an exception. We can observe that men were building compasses long before the *De Magnete* of Gilbert was published. And we must note also that this book, which is the first modern study on magnetism, did not influence the art of navigation. There was no change until the middle of the 19th century. Still at that time, steam engines were working for 70 years before someone tried to draw up their theory. And the building of machine-tools antedates by a good deal theoretical works of 19th century engineers.

Some French anthropologists or philosophers of life and technology have drawn drastic consequences from these observations. The most demonstrative argument in this way is the example of the locomotive. As Canguilhem¹⁶ says, the building of the steam engine remains unintelligible if one does not know that it is not by applying previous theoretical knowledge that it could be done but by finding a solution to a millenary problem, that of mining drainage. In the same order of ideas, Leroi-Gourhan goes further when he maintains that it is, in fact, the spinning wheel which stands at the origin of steam engines and today’s machines.¹⁷ So Canguilhem concludes that science and technology are two kinds of activities which cannot be grafted nor transplanted on one another, each one contenting itself with borrowing from the other, sometimes particular solutions, and sometimes particular problems. For Canguilhem, the origin of such a technology is an irrational one, and it is only because we are used to rationalize our techniques that we should forget this

irrational origin of machines. So we must make room for the irrational in the evolution of technology, even and especially if we want to support rationalism.¹⁸

The relationship of this thesis with that of Bergson one is obvious. In *Les deux sources de la morale et de la religion*, Bergson explicitly wrote that the spirit of mechanical inventions, though science keeps going on, remains rather distinct, and if need be, could part with it.¹⁹ For Bergson is one of the few French philosophers to consider mechanical invention as a biological function, an aspect of the organization of matter by life. But Leroi-Gourhan is very close to him when he tries to understand the phenomenon of tool-making by comparing it with the movement of an amoeba, pushing outside of its mass an excrescence which takes and captures the external object of its covetousness. On the contrary, for the Bachelardian school, evolution of life or evolution of technology will look much more like a broken line than a smooth slope.

4 The Technical Phenomenon

With Jean-Pierre Sérís, we must now come into the technical sphere and the world of technology. Here we have to raise some questions about the technical phenomenon, which needs not only a phenomenological description (or description of its appearance) but actually an *objective* description, which will not necessarily coincide with the immediate point of view of the users or agents.²⁰

From the user's point of view, a technique is first a means or set of means to achieve one's ends. Though integrated to our habits as the short way, even the "best one way" to do something, a technique, which is an arrangement of means and mediations, is in fact constituted by chains, networks and systems. As Sérís says, the technical chain is a path in a pre-existing network of available means. The technical apparatus that Bachelard tried to describe in physical science²¹ with the concepts of "apparatus consciousness" and "special determinism" may appear in technology as different kinds of systems. So, we may ask three questions about that systemic representation :

First, is there one system or many? Bertrand Gille thought there was only one, because many technological systems tend to make one: for him, technology is a set of consistent means at the different levels of every structure of all the sets and all the fields.²² It was also the opinion of Jacques Ellul, whose concept of a "technical system", one of the dominant factors of the occidental world, was in fact self-contained. But one may consider such views simplifications: generally speaking, historical reality is more complex and technology is rarely close to the state of equilibrium and other requirements of a "system". For instance, in the first part of the 19th century, we cannot say that there is a technological system based on the steam engine because steam power overtook hydroelectric only around 1864. Also the case of the compass and printing in ancient China are very well known: for different reasons, these major inventions did not lead, at the time, to a technological system. In the same way, control techniques were mastered by the engineers of Alexandria, but none of them was able to invent the steam engine. So we must be very careful if we want the concept of system to be useful in the domain of technology.

The second point is the question of the nature of the system itself. Bertrand Gille was thinking that its main feature was coherency, self-organization and self-regulation. On the contrary, Jacques Ellul supported the idea that a technological system is accurately lacking of them. It is a fact that optimization, organization and management tasks are less and less abandoned to some “invisible hand”, automatic regulation or chance. They are claimed to be technical tasks dependent on the responsibility or competence of technology itself. In this context, Gabor’s law (everything which is possible will be necessarily realized) is not a pessimist’s report but just an act by which we take into account the fact that the graph of techniques is a quasi-complete one, i.e. a graph where (almost) everything is in communication with (almost) everything. In fact, it is not quite true. In the domain of technology, exclusions and correlations may be so that every arrangement is not necessarily practicable.

In the end, technological systems are historical ones. What about their evolution? For some authors like M. Bloch, M. Mauss, A. Leroi-Gourhan, A. G. Haudricourt, B. Gille or M. Dumas, the idea of system specifies a domain of research (the “technical history of technology” from Lucien Febvre, for instance, or technology as a part of ethnography). And so, human reason, as in Hegel’s philosophy of History, may understand what happens. But for others (Jacques Ellul, Gilbert Hottois, Michel Serres or Michel Henry), the concept of a technological system essentially allows us to speak about the present time, especially for making a diagnosis of it, and a prognosis on its future. For the former, the technological system is an object which surely can be known; for the latter, it is an object which slips from our hands. These accusers of today’s technology denounce a breaking with the past. The autonomy of it is thought of as an infernal machinery left to a catastrophic and giddy progress, at any rate when things are following their usual train. And nobody can actually direct it. So a technological system is a dynamical one which appears, more and more, to be “out of control”. In fact, since the paleolithic Age, History shows many kinds of technological systems with a lot of periods of stagnation and revolution. There are many structural or extrinsic reasons which can explain why a technological system advances or halts. As Jean-Pierre Sérís showed it, jamming may have social or political reasons (China), ideological ones (Alexandria), religious ones (Muslim world after the year one thousand), or material ones (the lack of some raw materials, for instance). These are extrinsic reasons. But there are also structural ones. Haudricourt, for instance, explains that the lack of the wheel in pre-Columbian America is not due to the fact that the natives could not invent it. In fact, the invention of wheel can only take place in an agricultural civilization with a flat land, when men succeed to domesticate big herbivores. Without draught animals, one cannot get a continuous movement. Quick elimination of horses and elephants would have deprive Paleolithic hunters recently arrived on the continent of technical reasons to bring the wheel in. This kind of structural reason is not due to the system as a system. There are many other aspects of technology that French philosophy has well studied (normativity, historicity, relationship with machines, arts or responsibility) that we cannot develop here. For more information, the reader may be referred to Sérís (1994).

5 Machines, Engines, Automata, Networks

A history of the philosophical attitudes towards technology, and particularly towards machines, was written by Pierre Maxime Schuhl, who published, in the mid-20th century, a remarkable little book on the subject.²³ This history may be summed up the following way: it begins with a resignation without hope (Antiquity), and after an enthusiastic and very promising period (roughly, from 17th to 19th century), comes back to some hopeless feelings (contemporary thought). But the difference between the beginning and the end is obvious: it was due to the *absence* of the machine that the philosopher of Ancient times was sorrowing, whereas it is due to its *presence* that today's philosopher must resign himself.

These observations prove in fact that, very often, philosophers – especially French ones – are not interested in machine itself or in its technical reality, but in the machine as a human or social fact. In other words, the philosophical problem of mechanism does not depend on the place of the machine in production, but rather on its influence on human life.

Commenting on the book of Schuhl in some papers published in the journal *Critique* in the late 40s, Alexandre Koyré explained the failure of the Cartesian dream by saying that, in the succession of centuries, man realized that instead of becoming free and happy, he was going to be more and more a slave of the machine. Instead of being the Golden Age of humanity, the age of the machine was in fact her Iron Age. Needles and shuttles were now moving by themselves, as Aristotle, at the beginning of his *Politics*, wished it could be; but unfortunately, the weaver remained, today more than ever, chained to his work. French philosophy is full of lamentations of this sort. Proudhon, Fourier, Villermé and even Michelet, a supporter of mechanism, describing the 17 hours a day labor of the 19th-century workmen, all deplore the hard conditions and the extreme poverty following from it.

However, like Schuhl himself, and unlike Samuel Butler in *Erewhon*, Koyré does not proscribe machines, nor condemn them. He does not scrap anymore new manufactures or industries, and even disculpates the machine of the charge of necessitating continuous adjustments between man and his technical and cultural environment, leading in the end to a kind of Huxley's "brave new world" where, whatever the events may be, nobody protests against what happens. In fact, Koyré, in spite of all the negative aspects of the accumulative stage of capitalistic or even socialist systems, maintained that the technical intelligence of man has fulfilled its promise. Nothing is more characteristic of modern technology than the more and more widespread use of more and more artificial materials which are not to be found in nature as they are: alloy, glass, plastic materials,²⁴ now aramid fibers, Kevlar and kevlar reinforced composites materials, phenolic resins and glass-reinforced phenol, polyurethane engineering plastics, new thermoplastics, carbon and graphite fibers, polyester molding compounds, hybrid polyesters in general, new epoxy resin systems, etc.²⁵

So, trying to throw light on the reasons of the birth of technology, and of its stages of development, Koyré never gives up any right to naturalism nor to simplistic

explanations. At that time, two theories, in fact, were facing one another: the Marx-Engels theory of history supporting that, at a certain stage of their development, the material forces of production come into conflict with the existing relations of production, inducing a lot of change in social organization as a whole; the other one was the psychosociological explanation supported – among others (for instance, Emile Meyerson) – by Pierre-Maxime Schuhl. According to this philosopher, the technical stagnation of Antiquity, may be explained by the structure of ancient society and its economy. If machines were of no use, it was because Greeks or Romans had at their disposal living machines, cheap and numerous, as far removed from the free man as from the beast: slaves. This contrast between liberal and servile, prolonged by the difference between science and technology, would explain that one must wait until the end of Middle Ages and the growth of towns, trade, industry, to see the scientific method applied in the realm of practical experience. In other words, Schuhl was showing that, if the Ancient Greeks or Romans never developed machines (excepted clepsydra and some rudimentary mechanisms in Alexandria²⁶), it was because machines were something unimportant for them.

For Koyré, however, psychosociological theory does not provide a satisfactory answer. But it is not because another theory gives a better explanation. According to Koyré, we must realize that, in the history of technology, there is no general answer for everything. Perhaps we can indeed go on a bit beyond this prudent advice by making two points: first, the machines the engineers of Alexandria would have been able to build were steam engines. But steam engines need wood for working and there was not much wood left in Ancient Greece because of the excess of shipbuilding (triremes were big consumers of wood). Second, engines and factories lead to the development of what we can call mass production. But mass production must be sold off. And to this end, one needs not only some network (road, railways, and so on) but vehicles to run on it. In the Ancient Times, the only vehicle one can dispose of was the team (pulled by horses or oxes) and the rough nature of packsaddle forbids, in fact, too heavy loads. So machines were of no use at that time and this, probably suffices to explain why there were no engines in the Ancient Greece.

6 Technology and Ideology

Since the middle of the 1960s, a large part of the French tradition in philosophy became very critic towards technology, and sometimes said surprising things about it. Many people, who had never opened a science book nor gone into a factory or a laboratory, allowed themselves to pass judgments on things of which they had no knowledge of. Inspiration coming from Heidegger or Habermas took the place of thinking by themselves, and so, technology, from that time, tended to disappear behind a purely idealistic mode of thought.

Roland Barthes, in his *Mythologies*, showed that the objects of consumer society are surrounded by a halo of connotations in which they are captured as in a

net. These connotations which do not refer to Benjamin's aura nor to anything real, were supposed to make a system of signs which can be studied for itself. That is what Jean Baudrillard²⁷ did in *Le système des objets*, which developed a purely semiotic point of view about things becoming kinds of gadgets and being totally out of touch with the real world. Commenting on a striking image of the book (an iron, the bottom of which was covered with nails), Raymond Ruyer²⁸ made the common sense observation that such a thing, if it ever exists, could not be produced in a large quantity. A society rarely uses up its power and raw materials in vain.

Sfez position on the problem of health,²⁹ as Pierre Musso's on communication and networks, do not seem to me more tenable. After being very critical with decision theory, Sfez developed a critical approach of communication and has led to a denunciation of the ideology of perfect health. But who believes in that? Apart from a few American researchers who dreamed of a global ecological system equilibrium regulated, nobody thinks seriously that it is the main problem of the time. In the domain of medicine, crucial questions of the century are rather emerging diseases and the return, under another form, more virulent, of former infectious ones. The great challenge of our modernity is there, and nowhere else.

In the field of network philosophy, Musso³⁰ claimed that the notion of network, initiated by the Saint-Simonian school, deteriorated so to speak and rapidly transmuted into an ideology. For him, we find there a kind of networks' cult, at the origin of a theology of transparency which, for Philippe Breton,³¹ rather took place in the cybernetic project and the early writings of Norbert Wiener. Musso established a connection between this ancient cult of networks and more recent propaganda about information superhighways such as Bill Gates' view of the world.

But it does not seem to me that such considerations belong to philosophy of technology. The sociological and political approach of Sfez and Musso do not take into account the actual history of technological concepts nor theories. The authors only speak about what Hegel did not consider as concepts but as representations of concepts, which is not the same thing. And more than one philosopher today knows very well that the present time is more fertile in representations of concepts than in concepts.

7 Ethics and Technology

In a very paradoxical way, thinking about technology in France has often means to make some ethical observations on man and the relations between technology and morals. As it is said in a very typical book on the subject,³² technology is often viewed as a "social science". At least some scientists try to persuade themselves that it would have to be so. Why do, very often, thoughts about technology imperceptibly change into ethical observations? A possible answer is that, in the domain of technology, we are dealing with values. What is a technical value? For Séris,³³ it is a value of usefulness, which has to be distinguishable from market values

(not necessarily connected to the previous one), and from biological or aesthetic values. For instance, innovation is certainly a properly technical value, which has many consequences for society. Now we cannot say that this kind of value has absolutely nothing to do with aesthetics or biology, because a successful innovation may be a neat solution for a technical problem (which can lead, for instance, to some elegant object), and also because, in the end, this innovation, if it is useful, will be selected and will remain for a long time, possibly for ever, in the collective memory. So we are led to the following question: what is important for human societies, and not only at one point of their history but for the future? And this is an ethical question. Values of usefulness are inserted, in fact, in complex chains of operations which imply controversial debates, collective decisions and social commitments. For instance, ensuing public transport or developing solar power rather than nuclear power are technical choices linked to scales of values on which everyone does not necessarily agree. French philosophers often prefer such debates to serious and well-informed study of what technology, in concrete terms, is. So we have lost count of the malcontents and of their numerous bemoaning speeches: the imprecations of Janicaud³⁴ against the blind combinations of an autonomous technology for which everything is possible, the indignations of Virilio³⁵ against the collusion of technology and war, militarization of language or virtualization of the whole reality, the temptation of sinking into deep ecology, support of the so-called “rights of nature” in the papers of the disciples of Hans Jonas, the belief in the ideology of a “sustainable development”, popularized by the World Commission on Environment and Development (WCED) in its 1987 report entitled “Our common future”, as if classical technology aimed only at short-lived or ephemeral productions, and so on. But all this talking is not getting us anywhere and the progress of technology is sometimes a better contribution to peace than the dreams of UNESCO.³⁶ It is far from obvious that a dynamic balance between cultural differences and a supposed emerging global ethic is a key concept in educating for a sustainable future. Thinking that actions of people and businesses in their own communities, at local levels, may extend outwards in spirals of shared understandings and revised or renewed vision of things is an optimistic and probably completely idealistic point of view. This ideology, unfortunately, gains ground, while the critics of big technological systems,³⁷ forgetting the benefits they have brought to us, go on with their undermining. Another recent sign of the times is the importance devoted to technological risks in the context of nuclear sciences, genetic engineering or nanotechnologies. In a very dark and anxious book, Jean-Pierre Dupuy³⁸ shows that the frightening future of technology, which, according to him, is waiting for us, should be, in fact, a matter of public concern. Following H. Arendt, he recalls that technology, today, has the capacity of activating no return processes, and so, is more a matter of *action* than a matter of *production*. But whatever the importance of the danger (that we must not underestimate), the existence of possible technological disasters cannot forbid on their own the scientific and technological progress, especially since the former sources of power are coming to an end.

8 Technology and Metaphysics: The Ontological Approach

As I already said, a not inconsiderable part of French philosophy deals with the Heideggerian idea that science (and likewise technology) does not think anything. How is it possible to attach the least importance to such a vacuous statement is very mysterious. But a lot of French philosophers surely share with Heidegger the convoluted thought that “the question concerning technology is the question concerning the constellation in which revealing and concealing, in which the coming to presence of truth, comes to pass”.³⁹ What do they mean by that? It is hard to say and it is no more sure those bombastic words can hardly mean something, it would certainly be better not to waste one’s time to look for an answer.

However, as this kind of thought is now very widespread, we cannot avoid saying some words about it. The main purpose of Heidegger’s thesis is, in fact, to show that the sense of technology has changed with the development of science in the 17th century. From the earliest times until Plato, technology was no mere means. It was a kind of bringing-forth (*Her-vor-bringen* in German), a mode of revealing, linked to what we call truth, translating the Roman *veritas* and the Greek *aletheia*: “Technology is a mode of *aletheuein*. It reveals whatever does not bring itself forth and does not yet lie here before us, whatever can look and turn out now one way and now another”.⁴⁰ So, for the Greeks, in fact, technology and handcraft manufacture belonged to *poiesis*, as arts, poetics, but so does physics. Now when we say that modern technology is something incomparably different from all earlier technologies, it is not only because it is based on modern physics as an exact science. For Heidegger:

the revealing that holds sway throughout modern technology does not unfold into a bringing-forth in the sense of *poiesis*. The revealing that rules in modern technology is a challenging [*Herausfordern*], which puts to nature the unreasonable demand that it supply energy that can be extracted and stored as such.⁴¹

And while the work of the peasant does not challenge the soil of the field, mechanized industry is now challenging all the energies of nature. “Air is now set upon to yield nitrogen, the earth to yield ore, ore to yield uranium, for example; uranium is set upon to yield atomic energy”,⁴² and so on. A famous example is supposed to make us feel the difference between former and recent technology: the hydroelectric plant, set into the current of the Rhine, is not built into the Rhine River as was the old wooden bridge joining bank with bank for hundreds of years. This factory “sets the Rhine to supplying its hydraulic pressure, which then sets the turbine tuning. This turning sets those machines in motion whose thrust sets going the electric current for which the long-distance power station and its network of cables are set up to dispatch electricity”.⁴³ As Heidegger said, the river is now dammed up into the power plant, and what it is now, namely, a water power supplier, derives from the essence of the power station. For Heidegger, such a state of affairs is monstrous:

In order that we may even remotely consider the monstrousness that reigns here, let us ponder for a moment the contrast that speaks out of the two titles, “The Rhine” as dammed up into the power works, and “The Rhine” as uttered out of the art work, in Hölderlin’s hymn by that name.⁴⁴

The conclusion is that the Rhine is now reduced to “an object on call for inspection by a tour group ordered there by the vacation industry”.⁴⁵

Commenting on the different papers of Heidegger about the question of technology, Jean-Pierre Sérís showed that Heidegger’s ideas were not very new nor original ones. Jacques Ellul, in “La technique ou l’enjeu du siècle” (1954) and 20 years after in “Le système technicien” was already opposed to the current conception of modern technology which is criticized by Heidegger and may be called an “instrumentalistic anthropologism”. Gabor, Mumford, Illich, in the 1970s, were denouncing the supposed unlimited and so totalitarian modern system of technology. This leitmotiv of the 60s and 70s is still fashionable today. Bruno Latour took up the story around the 1980s, showing that scientific knowledge is, by role and necessity, totally unlimited, as capitalism which gave its birth to it.⁴⁶

Sérís himself was not convinced by Heidegger’s arguments though he understood very well its stake, which is to set out the Nietzschean “will for power” as a manner of fulfilling occidental metaphysics. But after the moment of criticism comes the question of knowing what we can do. And we cannot content ourselves with gazing at the marvelous old bridge of Heidelberg when extreme poverty, destitution, famine and diseases still exist in the world. Heideggerian condemnation of technology reveals in fact the casualness of a privileged being living in an opulent society, and who comes to forget what he is owing to it. I would also point out that an old bridge is again a technological object, which belongs to a network organization of the society certainly different from ours but historically specific and well defined. Perhaps it links together the church and the castle, or two trade fair towns on both sides of a river. But if we want to understand what technology is, we should not give priority to a particular state of its development.

9 Back to Technology Itself

Far from these critical and panoramic views, the Bachelardian school stressed the necessity to return to technology itself, and to begin to describe it.

Bachelard himself, who was first and foremost a philosopher of science, did not have a lot to say about technology. There are only a few passages devoted specifically to it in his works and most of them are dedicated to scientific apparatus. But there remain some in which we can disclose an actual interest in the technology of machines. For instance, we may read in *Le matérialisme rationnel*, that a machine is an inflatable apparatus which can be arranged in many ways according to its use.⁴⁷ And this observation has been very well verified afterwards when applied to machines like computers, which are governed by a program.

François Dagognet, a disciple of Bachelard born in 1924 and who wrote more than 50 books, is not only a philosopher of the life sciences. Throughout his work we find several reflections devoted to philosophy of technology. In *L’invention de notre monde, l’industrie: pourquoi et comment?* for instance Dagognet, after Marx, takes the factory as a philosophical object. For him, industry works won-

ders: from some cheap, ordinary and generally very abundant ingredients (the inputs) – that are modified, joined or transformed by some machines –, it makes valuable goods (the outputs). This process from less to more is something of a miracle. Though it is not a creation *ex nihilo*, the fact remains that it is a remarkable improvement.⁴⁸ Against all those philosophers who merely point out the negative effects of industrialization and content themselves with investigations based on too obvious charges, Dagognet, after Hume, Saint-Simon and Auguste Comte, says in manufacturers' defense that industry is actually a true demiurge. Against Marx and Marcuse, he shows that modern societies progressively create a set of laws which try to protect workmen as purchasers. Little by little, machines yield increasing outputs and more and more independence. The engines of Savery, Newcomen or Watt, lead to growing and growing productions. The perfecting of engines and the rules they must obey to give new forms of machinery and gears (Woolf, compounds, Ebinger machines). The progress of chemistry puts on the market new objects based on wood cellulose ($C_6H_{10}O_5)_n$ or polythene $(CH_2-CH_2)_n$ times, for instance, which gives polyvinyl chloride, PVC, when an atom of chlorine takes the place of an atom of hydrogen), and again polyamides like nylon and other polymers. By these inventions, industry does not only create a new universe, it changes the natural world into a human one, and establishes new kinds of relations between men or between men and objects. In fact, for Dagognet, industrial environment leads to a true religion and a burning art (*art brûlant*). In another book on the same subject, *L'essor technologique et l'idée de progrès*, Dagognet tries to save the idea of progress, even if, as he said, this idea may succumb under the blow of progress itself.

Though Dagognet wrote about many things, and also about technology, he is known, above all, as a philosopher of Chemistry and the life sciences. It was in fact another disciple of Bachelard, Gilbert Simondon (1924–1989), who developed the first great philosophy of technology of the second part of the 20th century. In a very dense book, entitled *Du mode d'existence des objets techniques*, Simondon, whose main problem was in fact the question of psychological and biological individuation,⁴⁹ tried to show how technology developed. In this book, he criticized Norbert Wiener's theory of cybernetics which, according to him, had accepted what any theory of technology must refuse namely a classification of technological objects conducted by means of established criteria and following genera and species. Simondon preferred to overcome the shortcomings of cybernetics and to develop a general phenomenology of machines, which shows especially, the numerous influences on the production of technical objects⁵⁰ and how, at every step of their development, these objects are synthetically reorganized – the engine, for instance, becoming more and more compact and its parts more and more interrelated.

After Simondon, Jean-Claude Beaune is surely, among the Bachelardians, one of the philosophers who has done the most to stimulate studies and reflections on philosophy of technology. Probably influenced by Mumford (1963, 1971), Beaune essentially took into account ambiguous and mythic entities. For instance, he wrote a beautiful book on the ambiguous concept of automaton,⁵¹ which is not

only a scientific notion but a philosophical one. The term of automaton, indeed, comes from Greek *automaton* which, in the beginning, meant nothing but chance. But the historical evolution of the word revealed another meaning, which is, on the contrary, completely deterministic: the *automaton*, as it is supposed to do always the same thing, is quite repetitious. Consequently, reason and unreason are involved in that concept, in an indissociable manner. We can even say more: with the great figure of repetition, which is linked to the Freudian principle of repetition, death is not very far off. Thus Beaune's observations on technology are also a meditation on death, and the relations between technology and death. One of the major themes of Beaune's philosophy is that man, when he tried to substitute technological systems for life – and that is one of the main and more useful aims of technology – never entirely succeeds because life is not comparable with a mechanical apparatus. When this confusion is made, then, the mechanization of life induces curious effects, like those we can see when medicine unduly prolongs human life, making of men some kinds of dead-living beings.⁵² Beaune also wrote about other ambiguous entities like remedies and drugs, and perhaps has been one of the first French philosopher to show that illness (or the scientific definition of the feeling of sickness) is, most of the time, a social disease: that is how former vagrants became, for 19th century's medicine, "itinerant automata", a short-lived and completely unfounded category.⁵³

In the last two decades, with the publication of four or five books, a new philosopher suddenly appeared in France, Gérard Chazal. This disciple of Dagognet, Beaune and Gayon, formulated first a philosophy of computer science,⁵⁴ before he developed it into a full-fledged philosophy of technology, a philosophy of forms⁵⁵ and even a theory of culture in general.⁵⁶ Influenced by the French mathematician René Thom (Fields Medal 1958), whose "catastrophe theory" belongs now to true science, Chazal, like Beaune, took an early interest in the concept of *automaton*. But he understood it essentially in the sense of computer science. For him, computer science allows us to go beyond logic and to get around the famous limits that Gödel's theorem (1931) impose on formal sciences. Also concerned with the concept of neuronal network, Chazal tried to use it to propose a new interpretation of the Aristotelician hylemorphic schema, which, according to him, may contribute to solve the mind-body problem by way of an updated materialism. In his latest books, Chazal attaches the utmost importance to the concept of "interface", which is, for him, not only the well-known computer science notion, but a true mediation which can explain many cultural facts or contemporary behaviors: from tattooing (because the body is an interface) to theory-making (when theories are some kind of abstract tools with which we can capture the world). In his latest book, *L'ordre humain ou le déni de nature* (2006), he has returned to the technology of building, information science as well as mind and body techniques, which are, for him, manners of making our (incomplete) world more human.

I have now a few words to say about my own involvement in the realm of technology. After working for more than 2 years with a space engineer at the CNES (Centre National d'Etudes Spatiales) in Toulouse, I wrote two books on modern

technology : *La conception technologique* (1998) and *L'homme volant* (2001). In these works, I tried to explain what modern design is, especially in the field of aeronautics and spacecraft engineering, and undertook to combat rampant “technophobia”, widespread in everyday life of all western societies. From the shipwreck of the Titanic to the nuclear accident of Tchernobyl, the comprehensible mistrust that could be generated by technological failures transmuted into an incomprehensible hatred based on ignorance and neglect of the great victories of humankind. I was naive enough to believe in a project of setting things straight, which was probably to bite off more than I could chew. But I maintain with Dagognet that French philosophers must leave their favorite meditations about consciousness and subjectivity, open their eyes if they are still prepared to look, and describe, if they are still capable of understanding what actually happened in the real world in the last century. The period 1900–2000 is not only the century of two bloody world wars. During those 100 years, man learned more about nature and its laws than during the whole stretch of earlier history. One of the main examples is the resolution of the problem of flying. From the dawn of aerodynamic thought to George Cayley, throughout the infancy of aerodynamics and its first applications by the Wright brothers, the scope of hydrodynamic phenomena subjected to exact analysis increased more and more, until a true science developed. Finally, the question of flying (or how to pose correctly the three problems of propulsion, lift and control) was to be solved by a meticulous study of gliding, even if the truth of the matter is that a powerful engine is enough to produce a propulsion and bring about a take off.

I want to conclude by saying that, like most of French philosophers, I am filled with admiration for the works of Jean-Pierre Sérís, whose premature death deprives French philosophy of technology of one of its best representatives. As former student of the *École normale supérieure* (rue d’Ulm), Jean-Pierre Sérís was evidently well versed in classical and modern philosophy. But he became also an expert in game theory and machine making, and probably wrote one of the best book published in France on the subject. Another of his books, on technology, entitled *La technique*, that I quoted often in this paper, is far more than a manual. It is one of the best guides I ever read on the topic, and should in my opinion remain for a long time a major book. I wanted to pay tribute to this modest, competent and clever man, Professor at the Sorbonne, Director of the Institut d’histoire des sciences et des techniques in Paris, a master and an example for all of us.

Endnotes

- 1 Descartes, 1953, p. 168.
- 2 Ibid., p. 566.
- 3 Descartes wrote a letter to Villebressieu in the summer of 1631, telling his friend that he carried out himself many experiments in mechanics. Cf. *ibid.*, pp. 943–945.
- 4 “A Huygens, 5 octobre 1637”, in *ibid.*, p. 971.
- 5 *Ibid.*, pp. 973–981.

- 6 Pascal 1954, p. 357.
- 7 See the conclusion of Dagognet, 1973. A recent disciple is Debray, 1991.
- 8 Poncelet, 1829.
- 9 Mitcham, 1994, p. 23.
- 10 Comte, 1830, pp. 66–69.
- 11 See Le Chatelier, 1925.
- 12 “Every change of one of the factors of an equilibrium occasions a rearrangement of the system in such a direction that the factor in question experiences a change in a sense opposite to the original change”, H. L. Le Chatelier, 1888, *Annales des Mines*, 13 (2), p. 157.
- 13 Bergson, 1970, p. 613.
- 14 Séris, 1994, p. 1.
- 15 Daumas, 1962, p. XI.
- 16 Canguilhem, 1967, p. 124.
- 17 Leroi-Gourhan, 1945, p. 100.
- 18 Canguilhem, 1967, p. 125.
- 19 Bergson, 1970, p. 1235 sq.
- 20 Séris, 1994, p. 46.
- 21 See Bachelard, 1951.
- 22 Gille, 1978, p. 19.
- 23 See Schuhl, 1947.
- 24 Koyré, 1981, p. 316.
- 25 See Parrochia, 1993.
- 26 See B. Gille, 1980.
- 27 Baudrillard, 1968.
- 28 See Ruyer, 1971.
- 29 See Sfez, 2002.
- 30 See Musso, 1997.
- 31 See Breton, 1987.
- 32 Bayle, Bourg et al., 1994, p. 51.
- 33 Séris 1994, p. 32 sq.
- 34 See Janicaud, 1985.
- 35 See Virilio, 1990, 1995.
- 36 I except the very good report of Ladrière (1977).
- 37 See Gras, 1993.
- 38 Dupuy, 2006, p. 135.
- 39 Heidegger, 1954 in Lovitt, 1977, p. 32.
- 40 *Ibid.*, p. 12.
- 41 *Ibid.*, p. 13.
- 42 *Ibid.*, p. 14.
- 43 *Ibid.*, p. 15.
- 44 *Ibid.*
- 45 *Ibid.*
- 46 Latour, 1988.
- 47 See Bachelard 1951, p. 13.
- 48 Dagognet, 1995, p. 24.
- 49 See Simondon, 1964.
- 50 Simondon, 1969, p. 72.
- 51 See Beaune, 1980a, b.
- 52 See Beaune, 1988.
- 53 See Beaune, 1983.
- 54 See Chazal, 1995.
- 55 See Chazal, 1997.
- 56 See the late developments of the philosophy of the author in Chazal, 2000, 2002, 2004 and 2006.

Bibliography

- Bachelard Gaston (1951), *L'activité rationaliste de la physique contemporaine*, Paris, PUF.
- Bachelard Gaston (1953), *Le matérialisme rationnel*, Paris, PUF, 3rd ed., 1972.
- Barthes Roland (1970), *Mythologies*, Paris, Point-Seuil.
- Baudrillard Jean (1968), *Le système des objets*, Paris, Gonthier-Médiations, 1975.
- Bayle F., D. Bourg, R. Debray, D. Ettigoffer, A. Finkielkraut, M.-A. Hermitte, B. Jacomy, D. Janicaud, B. Latour, R. Laufer, P. Lévy, A. Moles, J. Perrin, A. Picon, Ph. Quéau, M. Sicard, F. Sigaud, B. Stiegler, P. Virilio, J. Weissenbach (2000), interrogés par R. Scheps, Paris, Point-Seuil.
- Beaune Jean-Claude (1980a), *L'automate et ses mobiles*, Paris, Flammarion.
- Beaune Jean-Claude (1980b), *La technologie introuvable*, Paris, Vrin.
- Beaune Jean-Claude (1983), *Le vagabond et la machine*, Seyssel, Champ Vallon.
- Beaune Jean-Claude (1988), *Les spectres mécaniques*, Seyssel, Champ Vallon.
- Beaune Jean-Claude (1998), *Philosophie des milieux techniques*, Seyssel, Champ Vallon.
- Beaune Jean-Claude (2002), *Le balancier du monde*, Seyssel, Champ Vallon.
- Bergson Henri (1970), *Œuvres*, Paris, PUF.
- Blamont Jacques (1993), *Le chiffre et le songe, histoire politique de la découverte*, Paris, Odile Jacob.
- Breton Philippe (1987), *Histoire de l'informatique*, Paris, La Découverte.
- Breton Philippe, Rieu Alain-Marc, Tinland Frank (1990), *La techno-science en question*, Seyssel, Champ Vallon.
- Canguilhem Georges (1967), *La connaissance de la vie*, Paris, PUF.
- Carnot Sadi (1824), *Réflexions sur la puissance motrice du feu*, Paris, Blanchard, 1953, reed. Gabay, 1995.
- Chazal Gérard (1995), *Le miroir automate, Introduction à une philosophie de l'informatique*, Seyssel, Champ Vallon.
- Chazal Gérard (1997), *Formes, figures et réalité*, Seyssel, Champ Vallon.
- Chazal Gérard (2000), *Les réseaux du sens: de l'informatique aux neurosciences*, Seyssel, Champ Vallon, 2000.
- Chazal Gérard (2002), *Interfaces: enquêtes sur les mondes intermédiaires*, Seyssel, Champ Vallon.
- Chazal Gérard (2004), *Les médiations théoriques*, Seyssel, Champ Vallon.
- Chazal Gérard (2006), *L'ordre humain ou le déni de nature*, Seyssel, Champ Vallon.
- Comte Auguste (1830), *Cours de philosophie positive*, vol. 1, 1st lesson, Paris, Rouen frères (Bachelier); *The Positive Philosophy of August Comte*, transl. & condensation, H. Martineau, 1853.
- Condorcet Marquis de (1799), "Éloge de M. de Vaucanson", in *Éloges des académiciens de l'Académie royale des sciences*, vol. 3, Paris.
- Coriolis Gaspard Gustave de (1819), *Du calcul de l'effet des machines*, Paris, Carilian-Goeury, 1829.
- Couffignal Louis (1952), *Les machines à penser*, Paris, Minuit.
- Dagognet François (1973), *Écriture et iconographie*, Paris, Vrin.
- Dagognet François (1989), *Éloge de l'objet*, Paris, Vrin.
- Dagognet François (1995), *L'invention de notre monde, l'industrie : pourquoi et comment?*, Fougères, Encre Marine.
- Daumas Maurice (1962), *Histoire des techniques*, 5 vol., Paris, PUF, "Quadrige", 1996.
- Daumas Maurice (1953), *Les instruments scientifiques au XVIIe et au XVIIIe siècle*, Reed., Gabay, 2004.
- Debray Régis (1991), *Cours de Médiologie générale*, Paris, Gallimard.
- Descartes (1953), *Œuvres*, Paris, Gallimard.
- Dupuy Jean-Pierre (2006), *Retour de Tchernobyl, journal d'un homme en colère*, Paris, Seuil.
- Trans. J. Wilkinson, *The Technological Society*, New York, Alfred A. Knopf, 1964.
- Ellul Jacques (1954), *La technique ou l'enjeu du siècle*, Paris, Armand Colin, 1954.

- Ellul Jacques (1980), *The technological system*, New York, Continuum.
- Ellul Jacques (1990), *The technological bluff*, trans. Geoffrey W. Bromiley, Grand Rapids, Eerdmans.
- Friedmann Georges (1946), *Problèmes humains du machinisme industriels*, Reed, Paris, Gallimard, 1955.
- Gille Bertrand (1978a), *Histoire des techniques*, Paris, Gallimard, Encyclopédie de la Pléiade.
- Gille Bertrand (1978b), *Les ingénieurs de la Renaissance*, Paris, Seuil, "Point-Sciences".
- Gille Bertrand (1980), *Les mécaniciens grecs, la naissance de la technologie*, Paris, Seuil.
- Gras Alain (1993), *Grandeur et dépendance, sociologie des macro-systèmes techniques*, Paris, PUF.
- Guillerme André (1983), *Le temps de l'eau : la cité, l'eau et les techniques*, Seyssel, Champ Vallon.
- Habermas Jürgen (1968), *Technick und Wissenschaft als Ideologie*, Frankfurt am Main, Suhrkamp Verlag, 1968.
- Heidegger Martin (1954), "The Question Concerning Technology" in William Lovitt, *The Question Concerning Technology and Other Essays*, Harper & Row, 1977, pp. 3–35.
- Henry Michel, (1976), *Marx*, Paris, Gallimard, 1976.
- Henry Michel (1987), *La Barbarie*, Paris, Grasset.
- Hottois Gilbert (1984), *Le signe et la technique, la philosophie à l'épreuve de la technique*, Paris, Aubier.
- Hottois Gilbert (1996), *Entre symboles & technosciences, un itinéraire philosophique*, Seyssel, Champ Vallon.
- Jacomy Bruno (1990), *Une histoire des techniques*, Paris, Seuil.
- Janicaud Dominique (1985), *La puissance du rationnel*, Paris, Gallimard.
- Koyré Alexandre (1981), "Les philosophes et la machines", in *Études d'histoire de la pensée philosophique*, Paris, Gallimard, "Tel".
- Ladrière Jean (1977), *Les enjeux de la rationalité, le défi de la science et de la technologie aux cultures*, Unesco, Aubier-Montaigne.
- Laffitte Jacques (1932), *Réflexion sur la science des machines*, Reed, Paris, Vrin, 1972.
- Latil Pierre de (1953), *Introduction à la cybernétique, la pensée artificielle*, Paris, Gallimard.
- Latour Bruno (1988), "Le grand partage", *Revue du MAUSS*, 1 (3); pp. 2764.
- Latour Bruno (1992), *Aramis ou l'amour des techniques*, Paris, La Découverte.
- Lebeau André (1986), *L'espace en héritage*, Paris, Odile Jacob, 1986.
- Le Chatelier Henry Louis (1925), *L'industrie, la science et l'organisation au XXe siècle*, Paris, Dunod.
- Le Chatelier Henry Louis (1988), "Recherches expérimentales et théoriques sur les équilibres chimiques", *Annales des Mines*, 13(2), p. 157.
- Lecourt Dominique (1990), *Contre la peur*, Paris, PUF, "Quadrige", reed. 1999.
- Leroi-Gourhan André (1945), *Milieux et techniques*, Paris, Albin Michel, reed. 1973.
- Leroi-Gourhan André (1964), *Le geste et la parole*, vol. 1, *Technique et langage*, Paris, Albin Michel.
- Leroi-Gourhan André (1965), vol. 2, *La mémoire et les rythmes*, Paris, Albin Michel.
- Lévi-Strauss Claude (1962), *La pensée sauvage*, Paris, Plon.
- Lévi-Strauss Claude (1985), *La potière jalouse*, Paris, Plon.
- Lévy, Pierre (1987), *La machine univers*, Paris, La Découverte.
- Lévy, Pierre (1991), *L'idéographie dynamique, vers une imagination artificielle?* Paris, La Découverte.
- Mac Luhan, Marshall (1994), *Understanding Media*, Reed, MIT Press.
- Maitte Bertrand (1981), *La lumière*, Paris, Seuil, "Point-Sciences".
- Mauss Marcel (1948), *Sociologie et Anthropologie*, Paris, PUF.
- Mitcham Carl (1994), *Thinking through Technology, the Path between Engineering and Philosophy*, Chicago, University of Chicago Press.
- Moles Abraham (1970), *Créativité et méthodes d'innovation*, Paris, Fayard-Mame.
- Mumford Lewis (1963), *Technics and Civilization*, Harvest/HBJ Books.

- Mumford Lewis (1971), *Myth of the machine: Technics and Human development*, Harvest/HBJ books.
- Musso Pierre (1997), *Télécommunication et philosophie des réseaux, la postérité paradoxale de Saint-Simon*, Paris, PUF.
- Parrochia Daniel (1993), *Philosophie des réseaux*, Paris, PUF.
- Parrochia Daniel (1994), "A Historical Note on Artificial Life", *Acta Biotheoretica*, 43, pp. 177-193.
- Parrochia Daniel (1996) "Algorithmics and the Limits of Complexity", *Science in Context*, 9 (1), pp. 39-56.
- Parrochia Daniel (1998), *La conception technologique*, Paris, Hermès.
- Parrochia Daniel (2001), *L'homme volant, épistémologie de l'aéronautique et des systèmes de navigation*, Seyssel, Champ Vallon.
- Poncelet, Jean Victor (1829), *Introduction à la mécanique industrielle*, Paris, Gauthier-Villars, 1870.
- Ramunni Girolamo (1989), *La physique du calcul, histoire de l'ordinateur*, Paris, Hachette.
- Russo François (1955), *Eléments de bibliographie d'histoire des sciences et des techniques*, Paris, Hermann, 2nd ed., 1969.
- Russo François (1986), *Introduction à l'histoire des techniques*, Paris, Blanchard.
- Ruyer Raymond (1968), *La cybernétique et l'origine de l'information*, Paris, Flammarion.
- Ruyer Raymond (1971) *Les nuisances idéologiques*, Paris, Calmann-Lévy.
- Salomon Jean-Jacques (1984), *Prométhée empêtré. La résistance au changement technique*, Paris, Anthropos.
- Salomon Jean-Jacques (1986) *Les enjeux du changement technologique*, Paris, Economica.
- Salomon Jean-Jacques (1992), *Le destin technologique*, Paris, Balland.
- Schuhl Pierre-Maxime (1947), *Machinisme et philosophie*, Paris, PUF., 2nd ed.
- Séris Jean-Pierre (1974), *La théorie des jeux*, Paris, PUF.
- Séris Jean-Pierre (1987), *Machine et communication*, Paris, Vrin.
- Séris Jean-Pierre (1994), *La technique*, Paris, PUF.
- Serres Michel (1990), *Le contrat naturel*, Paris, Bourin.
- Sfez Lucien (1998), *La Santé Parfaite*, Paris, Seuil.
- Sfez Lucien (2002), *Technique et Idéologie*, Paris, PUF.
- Sigaut François (1994), "La technologie, une science humaine", in F. Bayle, D. Bourg, R. Debray et al, *L'empire des techniques*, Paris, Seuil, "Point-Sciences".
- Simondon Gilbert (1964), L'individu et sa genèse physico-biologique (l'individuation à la lumière des notions de forme et d'information), Paris PUF., 2nd ed. J. Millon, "Krisis", 1995.
- Simondon Gilbert (1969), *Du mode d'existence des objets techniques*, Paris, Aubier-Montaigne.
- Stiegler Bernard (1996), *La technique et le temps*, 2 vol., Paris, Galilée.
- Tinland Frank (1991) (ed.), *Systèmes naturels, systèmes artificiels*, Seyssel, Champ Vallon.
- Tinland Frank (1994) (ed.), *Ordre biologique, ordre technologique*, Seyssel, Champ Vallon.
- Virilio Paul (1977), *Vitesse et politique*, Paris, Galilée.
- Virilio Paul (1990), *L'inertie polaire*, Paris, Bourgois.
- Virilio Paul (1995), *La vitesse de libération*, Paris, Galilée.

Part II
General Philosophy of Science

A Problem in General Philosophy of Science: The Rational Criteria of Choice

Anastasios Brenner

1 Introduction

The question of the criteria involved in scientific choice is recurrent in philosophy of science. It has received attention in recent research and is not unrelated to several hotly debated issues: realism, truth, progress. The need to motivate decisions occurs in all areas of scientific inquiry. Thus the criteria of choice belong to a general philosophy of science, such as measurement or the structure of theories. By drawing attention to a question that cuts across disciplinary boundaries, we do not advocate the unity of science or a return to positivistic conceptions. We merely note that general questions arise, despite the fact that philosophy of science is branching off more and more into a series of distinct explorations of the various sciences. Those who adopt the disunity of science are not dispensed from having to explain interdisciplinarity. While new sciences emerge, each with its specific agenda, methods and techniques are transposed from one specialty to another.

Choice is an essential question in science as well as other areas of human activity. The criteria involved in choosing between theories help in turn to decide on a course of action in applied science. They may also provide insight for rational decision in everyday life. Yet it is far from obvious that the reasons we regularly give in favor of our choices have a clear, unambiguous meaning or that they have been submitted to a sufficiently thorough examination. Our aim in this article is to bring together the results of different strands of research: both logical analysis of established science and historical study of the development of science. Logical positivists as well as historically minded philosophers of science have paid heed to our problem. It is important to seek to determine the procedures deployed by contemporary science through an analysis of discourse and practice. But we should also endeavor to know where they came from and how they evolved, that is to describe their trajectories

A. Brenner (✉)
Département de philosophie
Université Paul Valéry
route de Mende
34199 Montpellier Cedex 5, France
E-mail: anastasios.brenner@wanadoo.fr

and variations. Logic does not dispense us from history, and we shall especially call on the historical approaches of Gaston Bachelard and Thomas Kuhn. To compare the views of these two major protagonists of the French and American schools of historical philosophy of science is not without interest in a volume devoted to French philosophy of science for English speakers. It will allow us to make some observations on the relation between these traditions.

2 The State of the Question in Kuhn's Second Doctrine

Let us take as our point of departure Kuhn's "Objectivity, Value Judgment, and Theory Choice". This article, included in *The Essential Tension* in 1977, deserves to attract our attention for several reasons. It provides the most complete presentation Kuhn has given of the criteria of scientific rationality. It addresses a major difficulty encountered by the doctrine of scientific paradigms. Furthermore, one may perceive here an evolution in the author's thought.

Kuhn had been accused of rendering scientific choice irrational in his 1962 *Structure of Scientific Revolutions*. Indeed, how to understand the transition from one paradigm to another that characterizes a scientific revolution? The paradigm being what explains the constitution of a particular community of researchers within society, it determines all at once the relevant problems, the appropriate methods and the shared values. In consequence, with the change of paradigm the scientific field concerned is subjected to a complete reorganization. There is thus no common basis for a strict comparison between two successive paradigms: one speaks then of incommensurability. Reflecting on this conception some fifteen years later, Kuhn endeavors to elaborate a new solution. He acknowledges that, although he had not until then scrutinized scientific values, he had not refused their existence.¹ One may nevertheless legitimately raise the question whether the introduction of this theme does not profoundly modify the manner of conceiving scientific change: the choice in favor of a new paradigm can now be rationally motivated and is founded on a series of recognized values.

Kuhn's answer consists in spelling out five basic requirements that should guide scientific decision: accuracy, consistency, scope, simplicity and fruitfulness. These values render possible a rigorous comparison of different theories. For instance, concerning simplicity Kuhn writes:

If [...] one asks about the amount of mathematical apparatus required to explain, not the detailed quantitative motions of the planets, but merely their gross qualitative features – then [...] Copernicus required only one circle per planet, Ptolemy two.²

The formulation suggests that, in historical context, the application of values is to be qualified. It is only with Kepler's elliptical orbits that heliocentricism became truly simpler. Kuhn proceeds to stress that different values may come into conflict; the final decision depends on the scientist's priorities:

Accuracy does permit discrimination, but not the sort that leads regularly to unequivocal choice. The oxygen theory, for example, was universally acknowledged to account for observed weight relations in chemical reactions, something the phlogiston theory had

previously scarcely attempted to do. But the phlogiston theory, unlike its rival, could account for the metal's being much more alike than the ores from which they were found.³

The history of science, which Kuhn calls on, shows that scientific victories are more complex than what rational reconstructions or manuals would lead us to believe.

What Kuhn is trying to make us understand is that the criteria of choice are not rules but values. None has precedence over the others, and there is no prescribed order of application. One must judge each case individually. Sometimes the choice between two theories will be made on grounds of simplicity, other times scope will serve as an argument. Kuhn does not believe in any automatic procedure of theory selection. There is no decisional algorithm. Kuhn thereby rejects both the inductivist techniques of the logical positivists as well as the falsificationist methods of Popper. We are concerned with a complex and subtle operation, which consists in interpreting and weighing the different criteria. Occasionally one must be ready to sacrifice one value to another. We may add a few more examples. Descartes, by restricting Aristotle's notion of motion to local transport, was able to formulate the principle of inertia; whereby he paved the way for modern physics. Maxwell, unlike Helmholtz, did not hesitate to juxtapose different models in order to produce an extremely fruitful theory of electromagnetism.

Rational criteria have a role to play in scientific activity. As Kuhn has it:

They do specify a great deal: what each scientist must consider in reaching a decision, what he may and may not consider relevant, and what he can legitimately be required to report as the basis for the choice he has made.⁴

According to Kuhn, criteria evolve more slowly than other theoretical components. Although communication among proponents of different theories may be difficult, they "can exhibit to each other [...] the concrete technical results achievable by those who practice within each theory. Little or no translation is required to apply at least some value criteria to those results".⁵

The solution offered in "Objectivity, Value Judgment, and Theory Choice" is not merely circumstantial. Kuhn will continue to concern himself with this theme in his posthumously published *The Road since Structure*.⁶ It is part of the new picture he gives of scientific change and marks, in my opinion, an important shift in his doctrine. Henceforth Kuhn draws a clear distinction between the adoption of a new paradigm by a scientist and its approval by the scientific community. The individual case is assimilated to a gestalt switch: the scientist suddenly sees the world differently. The phenomenon is obscure, and its analysis difficult. On the contrary, the collective case admits of a more direct examination. The acceptance of a paradigm by a group occurs gradually and gives rise to debates that witness the process of change. Kuhn mitigates the difference between normal science – the activity carried out within a paradigm – and extraordinary science – the search for a new paradigm. He no longer contrasts sharply periods of consensus with periods of disagreement. The constant aim of science is problem solving.

Kuhn also comes to emphasize language change. The conflict between Ptolemy's followers and those of Copernicus could be depicted as turning on the meaning of the word "planet": the Sun is no longer to be counted as a planet, whereas the Earth

becomes one. Such changes signal a discontinuity, but they only concern a small number of terms at the same time. What we are concerned with here are classifications, and Kuhn suggests that we direct attention to kinds. Classifications evolve, and their transformations can be studied. Now, rational criteria also give rise to classifications: kinds of activities represented by accuracy, consistency, simplicity etc. Together these values describe what we take for science. The meaning of the term science clearly varies more slowly than the individual theories upheld successively by scientists. Kuhn's amendments yield a more reasonable and intelligible view of scientific change. There is no doubt that he has provided some judicious qualifications. Several authors have followed up this line of thought and developed a careful examination of the manner in which we classify things.⁷

It remains that Kuhn's final doctrine raises several difficulties. He maintains that rational values change slowly. Now, this is a historical issue, and it is far from obvious that values cannot be affected by scientific revolutions. For example, before General Relativity it was taken for granted that Euclidian geometry provided the simplest and most convenient framework for physics. With Einstein's theory our understanding of simplicity was altered. Values, like the other components of the paradigm – symbolic generalizations, ontological models and methodological rules – may change during a revolution.

Kuhn is also rather vague about the meaning of the criteria of choice. Moreover, he does not tell us how he arrived at his aforementioned list of five standard requirements. Kuhn merely suggests that they are generally accepted. It is surprising that a historically oriented thinker should not have inquired into the origin of these criteria. Nowadays the anthropologist's rule to make explicit the standpoint from which a study is carried out has gained currency in all areas of the human sciences, and we require greater awareness on part of the philosopher of science with respect to the origins and nature of his concepts and methods.

3 Bachelard and the Transmutations of Rational Values

Concerning the values underlying scientific choice, Kuhn could have called on Bachelard. The latter had indeed directed his thoughts to this issue since his very first writings in the 1920s. Not only does Bachelard provide us with further observations on the question we are seeking to explicate, but he offers us a point of comparison. Reflection on rational values took different paths in France and America. Its history is made of strange detours, persistent misunderstandings and surprising convergences.

There are several striking analogies between Bachelard and Kuhn. Both thinkers react against the various forms of positivism upheld by their predecessors. Bachelard is critical of Comte and Mach; he even distances himself from the "new positivism" of Poincaré and Duhem. Kuhn opposes the logical positivism, then dominant in America. Bachelard and Kuhn seek to formulate a philosophy of science based on history of science. They defend a conception of scientific development in which

discontinuities occur locally and globally. Bachelard speaks of thresholds, mutations and epistemological breaks; Kuhn of conversions and gestalt switches. These underlie the revolutions that periodically shake science.

According to Bachelard, science goes through three stages: the pre-scientific mind, the scientific mind and the new scientific mind. The latter stage characterizes the recent transition from classical physics and chemistry to Relativity and Quanta. Kuhn similarly has preparadigmatic science, the paradigm of classical science and the new paradigm of the beginning of the 20th century. The Einsteinian revolution on examination shows that upheavals occur regularly in a discipline. The concepts of physics are profoundly modified, and a new worldview arises.

Bachelard brings out the tacit philosophy of scientists; the various aspects of scientific activity require resorting to different philosophical doctrines. Kuhn likewise reveals the metaphysical elements in scientific theories: the paradigm includes ontological models and rational values. The positivist ideal of philosophical neutrality is questioned. Historical study leads both thinkers to refuse simple philosophical models.

This parallel can be pursued with respect to rational values. Bachelard devotes ample space to them in his works. His doctoral dissertation, *Essai sur la connaissance approchée*, offers a detailed study of the concept of precision as it comes to play in contemporary science. He is opposing the conceptions of William James and Bergson. An analysis of the methods of approximation enables him to depart from these two philosophers. It is not enough to say that science aims to validate its hypotheses; for one must inquire into the degree of precision expected, that is the progressive application of methods of verification. Science does not remain at the surface of things. There is no need to contrast it with intuitive knowledge. This is what a careful examination of the concrete procedures of science shows: “Knowledge driven by methods of approximation will be able to follow the phenomenon right into its individuality and its very motion”.⁸ Bachelard thus rejects idealism and begins on his attempt to reformulate realism. He provides a series of observations that make it possible to reach a better understanding of the notion of precision. Bachelard distinguishes between a relative sense, *the precise*, and an absolute sense, *the exact*. Now, science is an endless succession of efforts to draw nearer to the object under study. Exactness with its connotations of certainty and perfection lead us astray:

As it is a pure impossibility to hit on the exact knowledge of some reality, even by chance – for a coincidence between thought and reality is a genuine epistemological monster – the mind must always summon up its capacities in order to reproduce the multiple diversities that designate the phenomenon under study by extending over its surroundings.⁹

Precision is neither complete nor final; it is approximate. The search for greater accuracy involves various factors: measurement, the application of mathematics to reality and the conditions of objectivity.

Measurement deserves to be considered first. Its operations can be translated without difficulty into the language of arithmetic, making possible a mathematization of the world. Measurement aims to go beyond the immediate perception of a property and to provide precision. With the remarkable development of instruments

and apparatuses in the 19th century, science came to be defined as the art of measuring exactly. Bachelard emphasizes the leaps and bounds made in precision; new orders of magnitude are brought into existence. But he is careful to adopt a critical attitude with respect to the traditional conceptions of Comte, Cournot and Maxwell: there is no question of an asymptotic approach to truth. Following on Mach, he reveals the philosophical assumptions that underlie measuring techniques. Progress in accuracy is to be correlated to the general advances of science.

But this is not enough for Bachelard. Scientific development is discontinuous. The increase in accuracy is not obtained by merely performing again the same operation with more care and rigor. One may need to change tactics, taking up a new hypothesis as guide or introducing an instrument based on another procedure. Measurement is not sheltered from the revolutions that frequently shake science. Bachelard stresses the dependence of measuring on the explanatory models and the experimental techniques. It is not sufficient to apply a unit of measurement to an object; the behavior of the instrument being used must be taken into account. One is inevitably led to formulate a theory of the instrument – which like any theory is susceptible of being continually improved. Accuracy is not given once and for all: it is inseparable from the general movement of science; it carries an irreducible historical dimension.

Bachelard does not fail to study how the realm of accuracy expands. Chemistry, for example, no longer contrasts the purity of a substance with its impurity; it takes into account degrees of purity as given by experimental techniques. Bachelard brings out the implications:

Other notions had their unity broken as soon as their precise conditions of application to the real were established experimentally. Thus the existence of pure bodies was once taken for granted in chemistry [...]. “Pure” is no longer for today’s chemist an adjective that does not admit of degrees [...]. The methods are what determine purity.

And we are given this warning:

But we may say that purity plays in matter the role of a Platonic idea in which the world participates; it is an ideal toward which the chemist tends by eliminating the impurities. One admits that he will never reach it. *We prefer to say that a meticulous chemist always reaches it.*¹⁰

The author breaks with the Platonic realism of mathematicians; he favors rather the idea of a realization.

Bachelard continues to explore rational values in his later works. We know that an important turn took place in his thought during the 1930s: he emancipated himself from his teachers, Brunschvicg and Abel Rey; his doctrine took on a firmer expression. In *The Philosophy of No*, he returns to the concept of precision:

No experimental result should be proclaimed as an absolute, divorced from the various experiments which have furnished it. *A precise result* ought even to be stated in terms of the various operations which produced it in its final form; operations which were at first imprecise, then improved, produced the established result. *No precision* is clearly stated without some history of initial imprecision.¹¹

Precision is never obtained at one stroke; it is the result of a gradual conquest. Bachelard now links this line of thought with the theses of his mature philosophy: the philosophical pluralism of science and phenomeno-technics.

More generally, Bachelard comes to speak of the “transmutations of epistemological values” and the “reaction of scientific knowledge on mental structure”.¹² It is the mind itself that changes in relation to scientific progress. He insists constantly on the various factors that enter into the elaboration of scientific knowledge: completeness, simplicity, consistency, accuracy. Bachelard is intent on drawing the full consequences of the revolutions that had shaken science and sketching the new scientific mind, as is expressed clearly in the conclusion to *The Philosophy of No*. According to him, non-Euclidian geometries ushered in the beginning of a profound intellectual mutation. It is not merely the question of their mathematical legitimacy but their relevance for physics, as became apparent with General Relativity. Quantum mechanics also leads to enlarge our conceptual framework. It is revealing that the proposal was made to modify classical logic based on the principle of bivalence. Bachelard underscores this direction by contemplating a possible transformation of arithmetic. Even if this discipline exemplifies rational values, it is not sheltered from scientific progress:

Arithmetic is not the natural outcome of an immutable reason any more than geometry is. Arithmetic is not founded upon reason. It is the doctrine of reason which is founded upon arithmetic [...]. In general the mind must adapt itself to the conditions of knowing [...]. The traditional doctrine of an absolute, unchanging reason is only one philosophy, and it is an obsolete philosophy.¹³

Bachelard is bringing classical rationalism into question and elaborating a dialectic rationalism. The succession of different theories over time should not lead to skepticism. One must be confident in the resilience of scientists. But the dialectical synthesis of earlier theories requires a recasting of rational categories. Dialectical overcoming and higher synthesis, such is the procedure proposed by Bachelard. But more to the point, this doctrine suggests that the criteria of choice have a philosophical content and are modified by scientific revolutions. In other words the criteria, like other aspects of scientific activity, are not unconnected with the experience of scientific life.

4 The Criteria in History

The preceding observations induce us to set about a more thorough historical analysis. Let us focus on the criterion of accuracy. The English language has constituted over time an array of terms to describe the mathematical rigor and meticulous observation characteristic of modern science: precision, exactness or accuracy. Our discussion being expressed in a particular natural language, we must be aware of the distinctions it suggests as well as those it dissimulates. These words derive from Latin, and their meanings have varied notably. “Exact” in its first sense designates the quality of what has been brought to perfection; “precise” what has been cut short; “accurate” what has been executed with care. They have acquired completely new meanings, and this evolution bears on the question at issue. Moreover, there are significant differences of expression among the major languages and cultures in which science developed; difficulties of translation may even arise.

The passage from Latin to the vernacular is often the occasion for a shift in meaning. Thus in translating Descartes' *Principia philosophiae*, we observe that the Abbé Picot, perhaps under the guidance of the author, substitutes the idea of exactness for that of uniqueness or univocality. The following passage is from the concluding section of the work; Descartes is arguing for the completeness of his explanation of natural phenomena:

Except for motion, magnitude and figure or situation of the parts of each body, which are things that I have explained as *exactly* [*exactement, unoquoque*] as possible, we perceive nothing outside of ourselves by means of the senses other than light, colors, odors, tastes, sounds and the qualities of touch: all of which I have proven are nothing apart from our thought but the motions, magnitudes and figures of various bodies.¹⁴

The exactness of the explanation is obtained by resorting strictly to geometrical notions. But this quality is not the foremost; it is rather "clarity" and "distinctness" that stand out as decisive attributes of his system. They make possible the mathematization of phenomena. But simultaneously they lead Descartes astray: the possibility of void is rejected on grounds of intelligibility.¹⁵

The term accuracy in the sense of "precision and correctness" enters the English language towards the middle of the 17th century and is used conspicuously in the context of the development of modern science. One of the earlier occurrences can be found significantly in Newton's preface to the first edition of the *Principia*: "accuracy" is employed here to characterize the distinctive trait of the science the author is expounding. It is not possible to offer here a full commentary. One would have to emphasize the rhetorical techniques that prepare the reader for the subsequent work. While calling on the authority of Pappus, Newton seeks to break with tradition. His discourse can be connected with the quarrel of the Ancients and the Moderns. But what is of interest for us here is the term accuracy, which occurs several times at the beginning of the preface. Newton proceeds to reverse the precedence of geometry over mechanics: "Geometry is founded in mechanical practice, and is nothing but that part of universal mechanics which accurately [*accurate*] proposes and demonstrates the art of measuring".¹⁶ The debate over the nature of the "exact sciences" probably originates here. This does not mean that earlier science did not seek rigor and certainty; Newton is of course promoting his theory. But undeniably new standards were being set, and these standards had an impact on our understanding of precision. Accuracy, as illustrated by Newtonian method, came to characterize science. The question remained how to understand scientific inquiry into the nature of man. For example Hume:

'Tis at least worth while to try if the science of man will not admit of the same *accuracy* which several parts of natural philosophy are susceptible of. There seems to be all the reasons in the world to imagine that it be carried to the greatest degree of *exactness*.¹⁷

The problem of the application of mathematics to the world also cropped up. Fernand Hallyn, in a study devoted to the rhetorical procedures of science, touches on the manner in which this debate was expressed during the 18th century. He observes how a series of thinkers – d'Alembert, Diderot and Buffon – were led to criticize the excesses of mathematization. We may retain the latter reflecting on the "union of mathematics and physics":

When I speak of the figures employed by Nature, I do not mean that they are necessarily or even *exactly* [exactement] similar to the geometrical figures that exist in our understanding; it is by assumption that we make them regular, and by abstraction that we make them simple. There are perhaps no *exact* [exacts] cubes nor perfect spheres in the universe [...]. The *precise* [précis], the absolute, the abstract, which present themselves so often to our minds, cannot be found in reality, because everything takes place by gradation, everything combines by approximation.¹⁸

In developing the contrast between human understanding and reality, Buffon brings to play several notions that deserve to be analyzed.

“Accuracy” is not the only criterion to have undergone transformation. We mentioned earlier “simplicity” in connection with Einstein. Let us now briefly examine the case of consistency. As used to signify the absence of contradiction of a mathematical system of propositions pertaining to natural phenomena, “consistency” came into currency during the late 19th century; the ancient equivalent “harmony”, as employed by Copernicus had a very different sense. In *The Copernican Revolution*, Kuhn brings out the neo-Platonism and the esthetic motivations that underlie the theory. His account differs curiously from the summary given later to the effect that simplicity is determinate.¹⁹ The question arises whether Kuhn does not concede too much to his critics in his later doctrine, and one wonders how to conciliate the historical study of the arguments actually put forth by scientists in favor of their theories at different periods and the rational analysis of criteria of choice in established science.

Rational values are not unrelated to our conception of truth. “Exactness” contains among its connotations the idea of an adequation between the thought and the thing. An exact theory could be understood as one that is in strict conformity with reality. This leads us to the conception of truth as correspondence. But nothing prevents us from calling on a new criterion. If we take the internal harmony of a theory as the index of its truth, we attach ourselves to the coherence conception of truth. To judge the theory on the basis of its predictions is rather to favor the pragmatist view. We encounter thus the major conceptions of truth. But there are in fact others, as suggested by our list of criteria. Simplicity could suggest a Platonic attitude: beyond the confusion of appearances one perceives the purity of archetypes. There remains the criterion of scope or, as it is sometimes formulated, completeness.²⁰ This requirement could serve, as in the passage of Descartes quoted above, not only as an argument in favor of preferring a particular theory but as a vindication of its truth. Although there is perhaps no major philosophical conception that one may refer to, completeness can be seen as related to the coherence conception and more particularly to holistic arguments.²¹ It is possible then to perceive metaphysical assumptions underlying the arguments employed in defending a theory. The variety of criteria available shows that scientific practice does not pledge allegiance to any particular conception. Scientific rationality, as Bachelard has it, implies a philosophical pluralism. Perhaps truth itself is protean.

After Newton the realm of accuracy expanded constantly from mechanics proper to numerous physical phenomena such as heat, electricity and magnetism. New areas came to be submitted to quantitative methods: chemistry, biology and sociology. In consequence, the question of distinguishing exact science from other forms

of scientific knowledge sprang up. Finally, as the Newtonian paradigm entered a state of crisis, a more reflexive, philosophical analysis of the nature of scientific practice arose. Mach criticized Newtonian notions of space, time and matter. He sought to relate these notions more precisely to the procedures of science. Accuracy began to take on its contemporary meaning: a precision of approximation.

5 The Debate over Theory Choice

The criteria as they evolved came to be associated with the problem of theory choice. This becomes apparent for example in Duhem's well-known definition of theory:

A physical theory is not an explanation. It is a system of mathematical propositions, deduced from a small number of principles, which aim to represent as *simply*, as *completely*, and as *exactly* as possible a set of experimental laws.²²

Simplicity, completeness and exactness explicitly enter into the definition of theory, and one needs only to read further to encounter the requirements of consistency and fruitfulness.²³ Thus we already find in *The Aim and Structure of Physical Theory* of 1906 the standard criteria evoked by Kuhn. This should not surprise us. Duhem is breaking away from the traditional view: a scientific theory is no longer conceived as an explanation of ultimate causes, but as an abstract representation of laws. One may perceive here the beginning of a conception that the logical positivists would amplify and that has been called the standard view: a theory is an axiomatic system – a set of propositions linked together by deductive rules; its empirical interpretation is provided by correspondence rules, comprising in particular procedures of measurements.

What is important to note is that “simplicity” does not refer to some ultimate, primitive quality, but to the provisional result of a procedure of decomposition. Likewise “exactness” is not to be given an ontological sense – the adequation between the thought and the thing. It should be understood in terms of a careful analysis of science:

The various consequences [...] drawn from the hypotheses may be translated into as many judgments bearing on the physical properties of bodies [...]. These judgments are compared with the experimental laws which the theory is intended to represent. If they agree with these laws to the degree of approximation corresponding to the measuring procedures employed, the theory has attained its goal.²⁴

Duhem was not alone in developing such ideas. He belonged to a movement of thought that included Poincaré, Le Roy and Milhaud, in which importance was given to conventions or decisions.²⁵

The novelty of this perspective can be brought out by a comparison with earlier conceptions. Let us return to Auguste Comte. He defined the positive or scientific spirit by means of the following series of terms: reality, usefulness, certainty and precision. He went on to add relativity as an afterthought.²⁶ The only one of our criteria he mentions is precision. These traits were supposed to define a new attitude. Comte was concerned with getting rid of the great metaphysical systems of the past

and elaborating a philosophy inspired by science. The question of the procedures of decision carried out by scientists was not raised. Comtian positivism provided a grand classification of the sciences and their progress, but paid little attention to the details of scientific practice.

Interestingly enough Bachelard singles out this passage for commentary:

It is perhaps the fourth characteristic [precision] that, as concerns measured phenomena, entails the others. Indeed the results of a measurement can sometimes be so *precise* that no account is taken of the very small errors that they still retain. These measurements, free from sensitive discrepancies, give rise unquestionably to a general consensus. It is by means of *precise* measurement that an object can reveal itself as permanent and fixed, in other words as a duly recognized object [...]. *Precision* takes precedence; it gives to *certainty* so solid a character that knowledge seems to us truly concrete and *useful*; it gives us the illusion of touching *the real*.²⁷

Bachelard is careful to leave aside positiveness, a term of which Comte made immoderate use. He emphasizes precision and shows how the other attributes derive from it. For sure, Comte had some interesting remarks to make: precision differs from certainty; it admits of degrees. This comes out in his presentation of the various meanings of the term “positive”:

A fourth commonly received meaning [...] consists in contrasting the *precise* with the vague: this sense evokes the constant tendency of genuine philosophical spirit to obtain in all areas *the degree of precision* compatible with the nature of the phenomena and in accordance with our true needs; whereas the old manner of philosophizing invariably led to vague opinions, which only achieved the necessary discipline through a permanent constraint based on supernatural authority.²⁸

Comte brought out the specificity of modern science and outlined the cognitive evolution of humanity. But his main concern was to avoid all metaphysics, going so far as to impose limits on science. Bachelard was to relinquish positivism. Science overturns obstacles; it renews the conditions of cognition and transforms rational values. In consequence, he could concentrate on the procedures whereby we establish precision.

We may now go over the history of the debate concerning theory choice. Kuhn borrowed his list of rational values in all likelihood from Hempel, with whom he was in close contact as colleague at Princeton shortly after the publication of *The Structure of Scientific Revolutions*.²⁹ Prior to Kuhn’s article on theory choice, Hempel devoted a chapter to the criteria of confirmation and acceptability in his *Introduction to Philosophy of Natural Science*.³⁰ He spelled out the conditions involved in theory confirmation as follows: quantity, variety, precision, predictions giving rise to new tests, simplicity and consistency. But more interesting still is Hempel’s article of 1983 “Valuation and Objectivity in Science”, in which he subjects Kuhn’s historical interpretation to critical scrutiny. He brings out certain analogies with Carnap and Popper, but disapproves of the pragmatist orientation. Hempel claims that Carnap had already taken into account several difficulties pointed out by Kuhn:

The problem of formulating norms for the critical appraisal of theories may be regarded as a modern outgrowth of the classical problem of induction: it concerns the choice between comprehension theories rather than the adoption of simple generalizations and the grounds on which such adoption might be justified.³¹

According to Hempel, Kuhn's conception does not make it possible to account for the rationality of scientific method. At the same time Hempel's criticism helps us understand how the account of theory choice evolved in relation to the general development of logical positivism. At an early stage logical positivists had set about to clarify a series of values that had come up as a result of earlier debates on scientific method. The standard view of theory structure was then gradually modified, and theoretical statements came to be conceived as making up strongly integrated systems whose relation to observation is complex and indirect. Hempel believes that these modifications suffice and that Kuhn's solution tends towards irrationalism or relativism.

6 Philosophical Semantics

It is time to draw attention to what we have been doing. The preceding study has sought to analyze how the criteria of choice function in scientific practice. Different meanings have been distinguished, and the variations over time have been outlined. One could characterize such an approach as a historical or philosophical semantics. We have concentrated on the criterion of accuracy, providing what should be considered only a rough sketch of the historical trajectory of the notion. Following these lines, a more thorough examination could be undertaken, especially with regard to the use of instruments, the correction of errors and the estimate of approximation. One could then move on to submit the other rational values to the same treatment, taking into account the full array of synonymous terms. What I should like to promote thereby is a research program concerning the process of decisionmaking in scientific activity.

Recourse has been made to both historical study and linguistic analysis, and we may call on the authority of several prominent thinkers. A recent trend seeks to reconcile the two major traditions in contemporary philosophy, the analytic and the historical, or at least to bring together their results. Paul Ricœur is a noteworthy example. Having studied in the phenomenological tradition, he came into close contact with analytic philosophy during the many years he spent teaching at the University of Chicago. In a late study on recognition, he associates significantly linguistic analysis and historical hermeneutics. Ricœur calls forcefully for associating both methods:

I have endeavored, as a good student of the good British school of ordinary language, to spell out the various meanings according to their particular context of use in natural language [...]. Before us the grand German philosophy of the 19th and 20th centuries had included philological inquiry in working out its guiding concepts. And, preceding us all, the Greek thinkers of the classical period, the good professor Aristotle foremost, pondered over the great book of customs as keen lexicographers, recording the appearance of suitable expressions in the poets and orators, before usage had erased the effect of these new additions to linguistic exchange.³²

Ricœur goes on to analyze the notion of recognition in this manner; he singles out three aspects: identification, reflexivity and mutuality. One could easily transpose these aspects to scientific knowledge. But what we should retain here is his

method. Linguistic analysis leads to historical considerations. Ricœur draws attention to various lexicographic tools that can provide the philosopher with relevant material. He interprets the linguistic facts, situates them within intellectual history and brings out the philosophical implications.

He is not alone in this venture. Jean Starobinski, in a study more closely related to science, *Action et réaction: aventures d'un couple*, follows in a similar manner the development of the concepts he has chosen. Both authors lead us to believe that from the point of view of practice the derivative precedes the root, the result is more important than the origin: thus recognition is prior to cognition; reaction to action. For we are already caught up in a series of events, in a chain of beliefs. This perspective leads Ricœur to a balanced treatment: recognition becomes the result of a complex effort involving not only the recognition of others but also self-recognition. As for Starobinski, he maps out the surroundings of scientific activity in terms that bring him near to our problem:

Objective knowledge which claims to consign consciousness to its origins (biological, neurological), and which seem to dispossess it must consent to recognize itself as the outcome of a decision, and bearer of future decisions.³³

But one can point also to an author seeking to bridge the gap from the other side: Ian Hacking endeavors to associate historical study and linguistic analysis. Although educated in the analytic tradition, he does not hesitate to call on French philosophy of science, the title of one of his books, *Historical Ontology*, is expressly borrowed from Michel Foucault. Hacking adopts it to characterize his own inquiry. Foucault had found inspiration in Bachelard and took his perspective further, making a broader and more systematic use of history, which he in due course named “archeology of knowledge” or “historical ontology”. Hacking takes up this approach, applying it more specifically to philosophy of science. He reformulates Foucault’s arguments: the concept of power is enlarged to include not only the power of repression but also the power of constitution. Thus is brought into view the interaction between the categories created by us and the objects or subjects subsumed under them. Finally, a concrete meaning is given to the attempt to relate discourse to its context of utterance. And Hacking offers a careful analysis of the production sites of experimental science: the laboratories and the research centers.

What is more, he claims that it is quite possible to recover thereby the concerns of analytic philosophy. Historical ontology is just another way of pursuing analysis: the conceptual usages are referred chronologically to their site of enunciation. Such inquiry can provide support in the quest to eliminate or overcome philosophical problems. Among the various currents of Anglo-American philosophy, some are not incompatible with the Bachelardian school; such is the case, according to Hacking, of ordinary language philosophy. This is how he presents his program:

Historical ontology is about the ways in which the possibilities for choice, and for being, arise in history. It is not to be practiced in terms of grand abstractions, but in terms of the explicit formulations in which we can constitute ourselves, formulations whose trajectories can be plotted as clearly as those of trauma or child development, or, at one remove, that can be traced more obscurely by larger organizing concepts such as objectivity or even facts themselves.³⁴

One can then submit the constitutive notions of science to a historical analysis, recording the discursive formulations and mapping out their development. The techniques Hacking employs betray similarities to those of Ricœur and Starobinski; historical ontology can be seen to merge with philosophical semantics.

7 Conclusion

One may perceive a shift in philosophy of science over the past century from the context of justification to that of discovery and invention. After the problems of induction and experimental testing interest has focused on the rational values attendant on theory choice. This leads to a richer model of scientific rationality. It could be feared that the historical perspective has the effect of undermining the authority of science: causes of belief are diverse, and motives change with time. History runs counter to the pedagogy of textbooks. This danger may threaten those accounts that remain polemic in attitude, aiming only to overthrow positivism. But at the same time history can provide its own antidote: it brings into view the actual procedures of science, those that have enabled the overcoming of obstacles on the path of discovery. Scientific rationality no longer appears static, formal, schematic, but rather intricate, flexible, self-correcting. Why not, with reference to the life sciences, speak of an adaptive, evolutive reason?

Furthermore, history reveals stabilizing techniques: the justification of a theory is progressively purged of its idiosyncratic elements. For example, the alleged harmony Copernicus claimed for his theory comes to be reckoned in terms of the logical consistency, the novel predictions and the theoretical fruitfulness of the heliocentric research program, its incompatibility with Aristotelian natural philosophy being resolved by the establishment of a new physics. With the development of science the grounds of belief are enhanced, multiplied and more and more closely coordinated. We need not commit ourselves to the unity of science to admit an encyclopedic task going on: scope and consistency push us continually to organize the fragments of our knowledge. The value of a theory is ensured by a whole network of reasons, which include both instruments and institutions. Above and beyond the empirical data dealt with is the full historical record: the difficulties overcome; the rival theories superceded.

Our investigation leads us to believe, against Kuhn, that the criteria of choice do not necessarily evolve slowly and do not transcend the other elements of the paradigm. Understood in a definite sense, they may undergo in time a profound transformation, and during a scientific revolution they may change suddenly. Rational values do not provide a way out of the difficulties of incommensurability. One must go beyond this framework. There is however an advantage to conceive these values as intimately bound up with the general movement of science. These variations are the outcome of discoveries. We are thereby more consistent with respect to the general tendency to deny strict separations between justification and discovery, fact and interpretation, theory and practice. Transitions in science appear in this light

perhaps more complex but also more intelligible. The criteria are given substance; they are related to concrete situations of decision.

Rational values call for a close analysis. They carry many different meanings. “Consistency” for example can designate any of a number of relations between two theories belonging to related areas of inquiry: entailment, reinforcement, compatibility, implausibility or inconsistency.³⁵ Our aim has not been to fix rigidly the meaning of the criteria; for we are not seeking to impose a formal algorithm of decision the difficulties of which have been stressed but to understand the paths, lengthy and varied, taken by scientific practice. The criteria must retain some flexibility or vagueness in order to fulfill their function: making possible a comparison of diverse paradigms and allowing a critical overview of the received paradigm. The terms employed are generally couched in natural language and play a meta-theoretical role. “Accuracy”, as we have seen, has a history of its own. Naïve conceptions of truth and reality have been left behind. What is now meant is a precision of approximation, resting on a whole array of techniques and procedures. Philosophical options have not been settled once and for all but enlarged; the scientist has acquired a larger compass for his research. Of course, precision must be accompanied by other advances, such as coherence and fruitfulness. The outcome is a global decision; the criteria summarize in a sense scientific activity. To devote attention to them is to take up the study of scientific knowledge, simply from a different angle.

Endnotes

- 1 See Kuhn, 1977, p. XXI. Indeed, the term value is rare in *The Structure of Scientific Revolutions*. It is in the postscript that Kuhn distinguishes four constitutive components of the paradigm: symbolic generalizations, models, exemplars and values.
- 2 Ibid., p. 324.
- 3 Ibid., p. 323.
- 4 Ibid., p. 331. Cf. p. 339.
- 5 Ibid., p. 339.
- 6 See in particular Chap. 9 “Rationality and Theory Choice”.
- 7 See for example Hacking, 1993, 2002.
- 8 Bachelard, 1928, p. 28. Translation mine.
- 9 Ibid., p. 43. Translation mine. Here, as in the following, I translate the French terms “*exact*” and “*précis*” systematically by “exact” and “precise”.
- 10 Bachelard, 1928, p. 80. Translation and emphasis mine.
- 11 Bachelard, 1940, p. 72. Emphasis mine.
- 12 Ibid., pp. 4, 7. Translation, pp. 5–7.
- 13 Ibid., pp. 144–145.
- 14 Descartes, 1644, vol. 3, p. 199.
- 15 Ibid., p. 203.
- 16 Newton, 1687, p. XVII.
- 17 Hume, 1740, pp. 645–646. Emphasis mine. Mill continues the debate: the realm of exactness is not defined once and for all. There are the exact sciences such as astronomy, the sciences in which exact laws are yet to be discovered such as meteorology and further removed but amenable to the same model the science of human nature. Cf. Mill, 1843, VI, Chap. 3, § 1, p. 844.

- 18 Buffon quoted in Hallyn, 2004, p. 209.
- 19 Kuhn, 1957, p. 172.
- 20 “Scope” is somewhat vague, and “completeness” expresses more clearly the aim of explaining comprehensively the phenomena belonging to a particular field. The latter should not be confused with the technical sense introduced by Gödel. “Completeness” is part of the traditional vocabulary of science. Thus Lagrange observes that Varignon’s *Nouvelle mécanique* “contains a complete theory of the equilibrium of forces in different machines”, *Mécanique analytique*, 1788, p. 13.
- 21 In particular, Duhem and Neurath place emphasis on completeness. They do not however base their theories on this sole criterion.
- 22 Duhem, 1906, p. 24. Translation, p. 19.
- 23 Duhem speaks of noncontradiction in the same passage. He later develops the requirement of consistency more fully, I, Chap. 4, §§ 7, 10; II, Chap. 7, § 1. Fruitfulness is examined further on, I, Chap. 3. It is a more complex concept, involving prediction and progress.
- 24 Duhem, 1906, p. 25. Translation, p. 20.
- 25 On this current of thought, see Brenner, 2003.
- 26 See Comte, 1884, pp. 120–126.
- 27 Bachelard, 1928, p. 52. Translation mine.
- 28 Comte, 1844, pp. 121–122. Cf. Comte, 1830–1842, I, pp. 60, 457.
- 29 Kuhn, 2000, p. 309.
- 30 Hempel, 1966, Chap. 4 “Criteria of Confirmation and Acceptability”.
- 31 Hempel, 1983, p. 92.
- 32 Ricœur, 2004, p. 13–14. Translation pp. 1–2.
- 33 Starobinski, 1999, p. 355.
- 34 Hacking, 2002, p. 23.
- 35 Cf. Laudan, 1977, p. 54.

Bibliography

- Bachelard Gaston (1928), *Essai sur la connaissance approchée*, Paris, Vrin, 1987.
- Bachelard Gaston (1934), *Le Nouvel esprit scientifique*, Paris, PUF, 1971; *The New Scientific Spirit*, trans. A. Goldhammer, Boston, Beacon Press, 1984.
- Bachelard Gaston (1938), *La Formation de l’esprit scientifique*, Paris, Vrin, 1975.
- Bachelard Gaston (1940), *La Philosophie du non*, Paris, PUF, 1994; *The Philosophy of No: A Philosophy of the New Scientific Spirit*, trans. G.C. Waterston, New York, Orion Press, 1968.
- Bachelard Gaston (1972), *L’Engagement rationaliste*, Paris, PUF.
- Black Max (1937), “Vagueness: An Exercise in Logical Analysis”, in *Language and Philosophy: Studies in Method*, Ithaca, Cornell University Press, 1949, pp. 23–58.
- Blackburn Simon (2005), *Truth: A Guide for the Perplexed*, London, Penguin.
- Brenner Anastasios (2003), *Les Origines françaises de la philosophie des sciences*, Paris, PUF.
- Brenner Anastasios (2006), “Quelle épistémologie historique? Kuhn, Feyerabend, Hacking et l’école bachelardienne”, *Revue de métaphysique et de morale*, 1, pp. 113–127.
- Comte Auguste (1844), *Discours sur l’esprit positif*, Paris, Vrin, 1995; *A Discourse on the Positive Spirit*, trans. E.S. Beesly, London, William Reeves, 1903.
- Comte Auguste (1830–1842), *Cours de philosophie positive*, 2 vols, Paris, Hermann, 1975; *The Positive Philosophy of August Comte*, trans. and condensation H. Martineau, 1853.
- Descartes René (1644), *Principia philosophiae; Les Principes de la philosophie*, in Adam and Tannery (eds.), *Œuvres*, 11 vols, Paris, Vrin, 1996.
- Descartes René (1644), *The Philosophical Writings of Descartes*, trans. J. Cottingham, R. Stoothoff, D. Murdoch and A. Kenny, 3 vols, Cambridge, Cambridge University Press, 1984–1991.

- Duhem Pierre (1906), *La Théorie physique, son objet, sa structure*, Paris, Vrin, 1981; *The Aim and Structure of Physical Theory*, trans. P.P. Wiener, Princeton, Princeton University Press, 1991.
- Hacking Ian (1993), "Working in a New World: The Taxonomic Solution", in Horwich, 1993.
- Hacking Ian (1999), *The Social Construction of What?* Cambridge, Massachusetts, Harvard University Press.
- Hacking Ian (2002), *Historical Ontology*, Cambridge, Massachusetts, Harvard University Press.
- Hallyn Fernand (2004), *Les Structures rhétoriques de la sciences: De Kepler à Maxwell*, Paris, Seuil.
- Hempel Carl (1966), *Philosophy of Natural Science*, Upper Saddle River, New Jersey, Prentice Hall.
- Hempel Carl (1983), "Valuation and Objectivity in Science", in *Physics, Philosophy and Psychoanalysis: Essays in Honor of Adolf Grünbaum*, R.S. Cohen and L. Laudan (eds.), Dordrecht, Reidel, pp. 73–100.
- Hume David (1740), *A Treatise of Human Nature*, Oxford, Clarendon Press, 1978.
- Horwich Paul (1993), *World Changes: Thomas Kuhn and the Nature of Science*, Cambridge, Massachusetts, MIT Press.
- Kuhn Thomas (1957), *The Copernican Revolution*, Cambridge, Massachusetts, Harvard University Press, 1977.
- Kuhn Thomas (1962), *The Structure of Scientific Revolutions*, Chicago, University of Chicago Press, 1970.
- Kuhn Thomas (1977), *The Essential Tension*, Chicago, University of Chicago Press.
- Kuhn Thomas (2000), *The Road Since Structure*, J. Conant et J. Haugeland (eds.), Chicago, University of Chicago Press.
- Laudan Larry (1977), *Progress and Its Problems*, Berkeley, University of California Press.
- Lagrange Joseph Louis (1788), *Mécanique analytique*, in *Œuvres*, vols 11, 12, Paris, Gauthier-Villars, 1881.
- Lagrange Joseph Louis (1788), *Analytical Mechanics*, trans. A. Boissonnade and V.N. Vagliente, Dordrecht, Kluwer, 1997.
- Mill John Stuart (1843), *A System of Logic*, in *Collected Works*, J.M. Robson (ed.), vol. 7, 8, London, Routledge, 1996.
- Newton Isaac (1687), *Philosophiæ naturalis principia mathematica*, A. Koyré and I.B. Cohen (eds.), 2 vols, Cambridge, Massachusetts, Harvard University Press, 1972.
- Newton Isaac (1687), *Mathematical Principles of Natural Philosophy*, trans. A. Motte and F. Cajori, 2 vols, Berkeley, University of California Press, 1962.
- Ricœur Paul (2004), *Parcours de la reconnaissance*, Paris, Albin Michel; *The Course of Recognition*, trans. D. Pellauer, Cambridge, Massachusetts, Harvard University Press, 2005.
- Starobinski Jean (1999), *Action et réaction: Vie et aventures d'un couple*, Paris, Seuil; *Action and Reaction: The Life and Adventures of a Couple*, trans. S. Hawkes and J. Fort, New York, Zone Books, 2003.

Science and Realism: The Legacy of Duhem and Meyerson in Contemporary American Philosophy of Science

Sandra Laugier

Anglo-American epistemology has long recognized its debt to Pierre Duhem: most notably in the so-called “Duhem–Quine” thesis that has been at the center of debates over empiricism and realism. These debates began with the Vienna Circle and have continued through the development of a more historical reflection on the sciences. This development is still ongoing, as can be seen in Hilary Putnam’s work on realism. The most prominent figures in this movement of inheritance of Duhem’s work, as well as the most controversial, are Kuhn and Feyerabend. But this change in American philosophy of science since, say, the sixties may also draw our attention to another influence, less visible than Duhem’s, but just as important: that of Emile Meyerson. One finds references to Meyerson in writings by both Quine and Kuhn. Kuhn, in particular, has explicitly recognized his debt to the author of *Identity and Reality*. In an interview in the French newspaper *Le Monde*,¹ he noted that he had, in philosophy, three major influences, apart from his contemporary, Quine: Duhem (for his *Aim and Structure of Physical Theory*), Meyerson (for *Identity and Reality*), and Koyré, who was responsible for the direct transmission of Meyerson’s work to the U.S. Kuhn also recalled that it was Popper himself who advised him to read *Identity and Reality*, a work that proved decisive for Kuhn.

These texts, somewhat forgotten in France after the thirties, were not only translated into English very early (*Identity and Reality* appeared in the U.S. in 1930), but were sometimes reprinted in English editions. In its original language, on the other hand, *Identity and Reality* has been unavailable for some time. In examining certain aspects of Meyerson’s work, we will attempt to understand why certain French philosophers, though forgotten in France until recently,² have been a source of inspiration for several philosophers of science in the United States.

S. Laugier (✉)
Département de philosophie
Université Picardie Jules Verne
Chemin du Thil
80025 Amiens, France
E-mail: sandra.laugier@u-picardie.fr

1 Holism and Ontology

This essay begins from two footnotes of Quine's. The footnotes in question come from an article that made a thunder-clap on the clear sky of analytic philosophy of science in the U.S.: "Two Dogmas of Empiricism", delivered as a lecture in 1950 and published in 1951. Quine, let us recall, was the one who, after a voyage to Europe in the 1930s, had introduced the work of Carnap and the Vienna Circle to American philosophers. In the wake of close collaboration between Quine and Carnap (as their recently published correspondence attests), a uniquely American form of logical positivism became the dominant movement in American philosophy departments. This movement was also encouraged by the forced immigration, in the thirties and forties, of many European philosophers and scientists, including Carnap, Reichenbach, Tarski, Frank, and Hempel. In his 1951 article Quine attacked the foundations of Viennese logical empiricism, namely the analytic/synthetic distinction and reductionism. These two dogmas are, according to Quine, "at root identical", and rest on a shared illusion: the possibility of distinguishing, in an utterance, between what belongs to experience and what belongs to language. In particular Quine singled out the distinctively neo-positivist idea that an utterance has an empirical meaning, and can as such be subject to empirical confirmation or refutation. Let us note that Quine was engaged here with an interpretation of logical empiricism that had become common, rather than with a serious reading of Carnap. It is too little noticed that Quine appeals to the *Aufbau* itself in his refutation of the second dogma.

My counter-suggestion, issuing essentially from Carnap's doctrine of the physical world in the *Aufbau*, is that our statements about the external world face the tribunal of sense experience not individually but only as a corporate body.³

But it is precisely here that we find a reference to Duhem, and not, as is often supposed, to the celebrated paragraphs⁴ on refutation and crucial experiment, but rather to the criticism of the Newtonian method.⁵

Quine does not so much take up the detail of Duhem's argument against refutation, but rather the general philosophy of *Physical Theory*, which one also finds in *To Save the Phenomena*, especially the impossibility of conceiving facts independent of all conceptualization. "An experiment in physics", Duhem writes, is:

quite another matter than the mere observation of a fact [...]. What the physicist states as the result of an experiment is not the recital of observed facts, but the interpretation and the transposing of these facts into the ideal, abstract, symbolic world created by the theories he regards as established.⁶

Their certainty, for Duhem, "always remains subordinated to the confidence inspired by a whole group of theories".⁷ It is precisely this point, taken up equally by Meyerson – "It is, as Duhem has justly said, impossible to understand the law, impossible to apply it, without performing the work of scientific abstraction, without knowing the theories which it presupposes"⁸ – that interests Quine. The testimony of experience, independent of any theoretical context, is a philosophical myth: "Statements, apart from an occasional collectors' item for epistemologists, are connected only deviously with experience".⁹ The critique of refutation, and what

we call “Duhem’s problem”, are a methodological consequence of this philosophical position, a position also adopted by Meyerson.¹⁰ No statement is in itself refutable, because there is no statement that speaks purely of experience (as the protocol sentences of Carnap’s *Aufbau* were supposed to do): even statements of experience are *theory-laden*. A recalcitrant experience therefore does not suffice to refute a theory: refutation is not as simple a matter as we might have thought.¹¹

Duhem’s idea that a negative experience does not require the rejection of a theory is frequently taken up by the post-Popperians. It was also developed in Quine’s philosophy of science in the form of his much-discussed epistemological holism. We find a very explicit formulation of Quine’s position in the introduction to his textbook *Methods of Logic*:

Statements close to experience and seemingly verified by the appropriate experiences may occasionally be given up, even by pleading hallucination”¹² This exactly parallels Duhem’s remarks: “When the experiment is in disagreement with [the experimenter’s] predictions, what he learns is that at least one of [his] hypotheses [...] is unacceptable and ought to be modified; but the experiment does not designate which one should be changed.”¹³

We can always preserve of a statement *come what may*, Quine concludes. On the other hand – an idea new with Quine – there are no unrevisable statements. Such are the interconnections assured by the logical relations between statements that every statement, even one taken as “central”, is vulnerable to a negative experience. Experience can have consequences anywhere in the system. “Reevaluation of some statements entails reevaluation of others, because of their logical interconnections – the logical laws being in turn simply certain further statements of the system, certain further elements of the field”.¹⁴ There is thus no privileged place within the conceptual scheme. Any statement, even one occupying a central place in the system, can be put into question. This is so even for logical laws, which, despite their “decisive position”, can be revised, if the revision provides a simplification necessary to the survival of the system.

Here again one can cite Duhem: the apparently unchangeable and necessary principles of physics, even those that cannot be directly subject to experiment, can be overturned in the development of science.

On that day some one of our hypotheses, which taken in isolation defied direct experimental refutation, will crumble with the system it supported under the weight of the contradictions inflicted by reality on the consequences of this system taken as a whole.¹⁵

Every statement is thus revisable. This is the meaning of the metaphor, favored by Quine and made famous by him, of a “field of forces” representing “the totality of science”,¹⁶ where statements confront experience at the periphery yet redistribute consequences to the interior, even to the most distant statements. There is no break between the periphery and the center, only differences of degree of proximity to experience, always provisional and never measurable: this is precisely the point that signaled Quine’s break with the Vienna Circle.

Thus we see that holism is, in Quine, a double-edged sword. Any statement can be revised, but, on the other hand, it is equally true that any statement can be preserved. On this point we can cite another passage from *Methods of Logic*:

Our system of statements has such a thick cushion of indeterminacy, in relation to experience, that vast domains of law can easily be held immune to revision on principle. We can always turn to other quarters of the system when revisions are called for by unexpected experiences.¹⁷

For Quine this holds not just for physics, but also for logic (though fundamentally revisable, it can be held immune “on principle”, because of its central place and also because of the indeterminacy of translation). This unnoticed consequence is nothing but the flip side of holism, or as Quine calls it, the “logical (rather than epistemological) point of view” on holism. For Quine there is no contradiction in this: the revisability of logic goes along with its immunity – these are just two sides of the same coin.

When some revision of our system of statements is called for, we prefer, other things being equal, a revision which disturbs the system least [...] despite the apparent opposition between this priority and the one previously noted, the one involves the other.¹⁸

It is because a revision is never merely local, but always “systematic”, that any revision must reflect choices and decisions according to what Quine calls “priorities”. There is no sense in revising the system unless one keeps it open, at each intermediate stage in the history of science, to revisions that will ensure its survival. “Mathematics and logic, central as they are to our conceptual scheme, tend to be accorded such immunity, in view of our conservative preference for revisions which disturb the system least”.¹⁹ Nonetheless there are priorities and conditions that decide the place of a hypothesis in the system. Briefly, we choose on a pragmatic basis²⁰ the change that disturbs the system least, unless a more wide-ranging revision offers other advantages, in particular *simplification*.

It is with reference to Duhem that Quine draws his most “anti-realist” conclusion: physical theory is not an explanation, but a symbolic representation: after recalling Neurath’s metaphor of the boat (the philosopher is “a mariner who must rebuild his ship on the open sea”), he adds:

We can improve our conceptual scheme, our philosophy bit by bit, while continuing to depend on it for support; but we cannot detach ourselves from it and compare it with an unconceptualized reality. Hence it is meaningless, I suggest, to inquire into the absolute correctness of a conceptual scheme as a mirror of reality. Our standard for appraising basic changes of conceptual scheme must be, not a realistic standard of correspondence to reality, but a pragmatic standard.²¹

Quine concludes with an appeal to “conceptual economy” that harkens back to both Duhem and Mach, recalling also the “pragmatist” tone of “Two Dogmas”.²² The first concern of holism is conservatism, or, to put it more naturalistically, the survival of the conceptual scheme. Transformations of the system, even radical ones, are gradual. Conceptual change, even major change, can be effected without a sharp break. It is simply because a revision is never merely local, but always systematic, that choices must be made. There is no sense in revising the system unless one keeps it open, at each stage in the development of science, to revisions that will maintain its stability. This is also the meaning of Neurath’s metaphor: “Our boat stays afloat because at each alteration we keep the bulk of it intact as a going concern”.²³

This Quinean model of the development of science, at once conservative and revolutionary, was in fact sketched out in a metaphor that Duhem uses in *Physical*

Theory: “Physical science is a system that must be taken as a whole; it is an organism in which one part cannot be made to function except when the parts that are most remote from it are called into play, some more so than others, but all to some degree”.²⁴ And it is remarkable that Duhem, in order to illustrate the difficulty of refutation, uses a biological metaphor: a physicist cannot determine the exact place at which his theory has broken down, just as a doctor

has to guess the seat and cause of the ailment solely by inspecting disorders affecting the whole body [...] The watchmaker to whom you give a watch that has stopped separates all the wheelworks and examines them one by one until he finds the part that is defective or broken [...] Now the physicist concerned with remedying a limping theory resembles the doctor and not the watchmaker.²⁵

This metaphor shows that Duhem’s doctrine of continuity is based on a form of epistemological holism that finds its most developed expression in Quine.

From this point of view, if we now return to “Two Dogmas”, we will not be surprised to find reference in the text to another French philosopher – Meyerson. Epistemological holism, the impossibility affirmed by Quine of determining the adequacy of our conceptual scheme as a representation of reality, seems little compatible with the frankly ontological philosophy of *Identity and Reality*. How can we, at this point in the discussion, invoke ontology? We might recall that Duhem did not rule out an ontological order:

Thus, physical theory never gives us the explanation of experimental laws; it never reveals realities hiding under the sensible appearances; but the more complete it becomes, the more we apprehend that the logical order in which theory orders experimental laws is the reflection of an ontological order.²⁶

There is indeed a form of realism in Duhem, in his idea that “these theories are not a purely artificial system, but a natural classification”.

But it is clear that this is not Quine’s position. The idea of “realities hiding under the sensible appearances” is quite far from his approach, precisely because of his interpretation of the ontological problem. Toward the end of “On What There Is”, Quine suggests that from a phenomenalist point of view, ontologies that include physical or mathematical objects are “myths”.²⁷ This notion of “myth”, which comes back later in “Two Dogmas”, has reinforced conventionalist interpretations of Quine: “The myth of physical objects is epistemologically superior to most in that it has proved more efficacious than other myths as a device for working a manageable structure into the flux of experience”.²⁸ Quine famously compared the ontology of physics (not just that of objects, but also, e.g., that of forces – a topic dear to Meyerson) to that of the Homeric gods. There is in this a clearly instrumentalist conception of ontology (this can be traced back to Mach): “physical objects are conceptually imported into the situation as convenient intermediaries, not by definition in terms of experience, but simply as irreducible posits, comparable, epistemologically, to the gods of Homer”.²⁹

We might wonder, then, how the empiricism of “Two Dogmas” amounts to a criticism of neo-positivist epistemology, since the neo-positivists more or less adopted wholesale Duhem’s idea that physical theory was a symbolic representation and a formal system. We can see this in Carnap’s *Logical Syntax*: “it is, in general,

impossible to test even a single hypothetical sentence [...] Thus *the test applies, at bottom, not to a single hypothesis but to the whole system of physics as a system of hypotheses*".³⁰ Clearly what is at stake here is not just holism, but realism as well. Quine's references to Meyerson is not merely out of respect or empty, and perhaps Quine is more serious than we suppose when he appropriates the statement from *Identity and Reality* "L'ontologie fait corps avec la science elle même et ne peut en être séparée ["Ontology is part of the body of science itself and cannot be separated from it"]."³¹ Let us look at the context in which this reference appears more closely. Quine affirms the continuity of ontological and natural scientific questions in the essay "On What There Is". The problem of ontology, according to Quine, is not that of knowing what exists, but of knowing the ontological significance of our discourse – of knowing what *we say* exists. Ontology does not, therefore, have for Quine the task of determining what there is. "What is under consideration is not the ontological state of affairs, but the ontological commitments of a discourse".³² The ontological question is transformed: "But we have moved now to the question of checking not on existence, but on imputations of existence: on what a theory *says exists*".³³ In order to know "what exists", one must look not to ontology, but to science. What exists is what science, as a whole, "says exists". And just as the only possible response to the ontological question is within science, the philosophy of science is identified by Quine with ontology.

One might, to parody a phrase of Wittgenstein's, say that "it is science (not grammar) that tells us what sort of thing something is". The citation from Meyerson therefore announces Quine's naturalism well before "Epistemology Naturalized". According to Quine's naturalism there is no fundamental difference between the task of philosophy and that of science. Ontology is an enlargement and generalization of scientific achievements. On the other hand, Quine suggests, in "Things and their place in Theories" that epistemology is a "methodology of ontology". The work of ontology is no different from the work of science, and participates in the same process of continual systematic revision. The philosopher's task is that of

making explicit what had been tacit, and precise what had been vague; of exposing and resolving paradoxes, smoothing kinks, lopping off vestigial growths, clearing ontological slums. The philosopher's task differs from the others', then, in detail; but in no such drastic way as those suppose who imagine for the philosopher a vantage point outside the conceptual scheme that he takes in charge.³⁴

There are no more privileged objects than there is a privileged science; there is a continuity, from the middle-size objects whose names we learn in first learning language, to the most sophisticated objects of science. "All objects are theoretical". For Quine, the ontology of science, even when it posits objects quite distant from our experience, is an extension of the ontology of common sense, not because the latter is in some way supreme, but because it is already theorized.³⁵

Here we find again a connection with Meyerson. The work of science, even in the context of naturalized epistemology, is ontological: as Meyerson had already said, science does not content itself with establishing laws. "Whatever opinion or system one supposes to prevail from a strictly philosophical point of view, one must admit that science itself is and remains a creator of ontologies".³⁶

Beginning in *Identity and Reality*, Meyerson affirms that

the ontological character of scientific explication is ineffaceable [...] there is not, there cannot be, in the natural evolution of scientific theories, any phase where ontological reality would disappear, and at the same time the concept of conformity to law remain standing.³⁷

Meyerson was the first to propose a model of the evolution of science in which ontological change defines scientific change (even if it is governed, as always for Meyerson, by the principle of identity). Moreover, it is these changes of ontology that allow him to describe, in *Identity and Reality* and *Explication in the Sciences*, the conceptual changes made in the history of the sciences. These changes are always motivated by the emergence of a new ontology: “the scientific intellect imperiously demands an ontological reality, and if science did not permit the creation of a new one, it would certainly be powerless to destroy the old one.”³⁸

There is no ontology independent of or prior to science: from this point of view, Meyerson is paradoxically less of a metaphysician than Duhem, and he prefigures Quine, even if Quine’s ontology relativizes and radicalizes Meyerson’s. (Quine proposes, in *Theories and Things*, that any ontology can be reinterpreted in the terms of any other via a “proxy function”.) The Meyersonian conception of ontology allows Quine, beginning in 1951, to make ontology immanent. This leads in his work to a final dissolution of the “question of transcendence” – that of the adequacy of physical theory to reality, or, as he put it in 1981, “the question whether or in how far our science measures up to the *Ding an sich*.”³⁹ On this point, Quine is far from Meyerson. But it is from Meyerson that he takes the idea of an immanent ontology, which is central for his work beginning with “Two Dogmas”. And one might say that it is over this point – realism – and not over holism that he breaks with Carnap.

For Quine, it is not possible that the philosopher take up “a vantage point outside the conceptual scheme that he takes in charge.” “There is no such cosmic exile”, he concludes in *Word and Object*. Carnap had already said as much in his *Logical Syntax*. But for Quine, ontology, once relativized, is not taken less seriously, and ontological relativity is in no way the dissolution of ontological questions. For Carnap, the question of a theory’s ontology is not a theoretical question, but a question that calls for a practical decision about the structure of our language. For Quine, by contrast, the question is more complicated because on his view “our theory of nature grades off from the most concrete fact to speculations about the curvature of space-time [...]. Existential quantifications of the philosophical sort belong to the same inclusive theory”.⁴⁰ General ontological questions, for Quine, are not a matter of language, or of the choice of a “conceptual scheme”, any more than ordinary scientific hypotheses are. The essential disagreement between Quine and Carnap is over ontology. Quine recognizes this at the beginning of his essay “On Carnap’s Views on Ontology”, which was a response to Carnap’s article “Empiricism, Semantics, and Ontology”. And we have suggested that the break represented by “Two Dogmas” comes not over epistemology but over the status of ontology. In sum, for Quine there is a continuity between talk about experience and talk about things, and ontology cannot be a matter of linguistic decision.

From this point of view, the appeal to Meyerson in *From a Logical Point of View* is paradoxically appropriate, even if Meyerson seems to postulate an independent reality from which science extracts or reconstructs its elements. Meyerson’s realism

requires that we take the ontology of science “seriously”, while at the same time taking account of the changes of ontology that have occurred in the history of the sciences. This is the only ontology we have at our disposal.

It is because Quine’s naturalism shares this approach to ontology that it does not exclude realism, even a “robust realism”, as he says in *Theories and Things*. Naturalism is “the recognition that it is within science itself, and not in some prior philosophy, that reality is to be identified and described”;⁴¹ it is the “abandonment of the goal of a first philosophy prior to natural science.”⁴² Even if we do not know whether our theory of the world or our ontology is the best or the only one possible, we must take it to be true. “We continue to take seriously our own particular aggregate science, our own particular world-theory or loose total fabric of quasi-theories, whatever it may be.”⁴³ Truth is *immanent*, and questions of *reality* cannot be posed except from within our system of the world. “There is no extra-theoretical truth”. This obviously poses a problem, which we will attempt to clarify. Quine’s naturalism, since it incorporates ontological questions into natural science, is a specific form of naturalism (irreducible, among others, to its cognitivist successors and by-products). Its ontological and realistic “theses” (found in the essays “Speaking of Objects” and “Ontological Relativity”) are inseparable from a radical skepticism about the possibility of determining a “natural”, preconceptual ontology, even by the most refined scientific methods. In a more recent text, Quine writes:

These reflections on ontology are a salutary reminder that the ultimate data of science are limited to our neural intake, and that the very notion of object, concrete or abstract, is of our own making, along with the rest of natural science and mathematics. It is our overwhelmingly ingenious apparatus for systematizing [...] our intake, and we may take pride.⁴⁴

Ontology is “a human option”, and the notion of reality “is itself part of the apparatus; and sticks, stones, atoms, quarks, numbers, and classes are all utterly real denizens of an ultimate real world, except insofar as our present science may prove false on further testing.”⁴⁵

Is there the same notion of ontology at work in this radical naturalism as there was in the quotation from Meyerson in “Two Dogmas”? Is there, in Meyerson’s philosophy of science, the possibility, taken up by Quine, of an ontology immanent in science? The question remains to be posed. But it was probably this idea of Meyerson’s, along with his reading of Duhem, that led to the decisive shift in Quine’s philosophy and that gave direction to his break with the “dogmas” of logical empiricism, i.e., with those of classical analytic philosophy of science.

2 Philosophy of Science and Realisms

We often, and rightly, consider the shift in American philosophy of science during the years 1960–1980 (initiated by Kuhn, Lakatos, Feyerabend, Putnam, and Hacking) as a break with mainstream philosophy of science based in logical positivism. This shift in philosophy of science had two aspects: a radically new conception of the *nature of science*, and an equally new approach to resolving the problems of

philosophy of science – an approach joined, inevitably, to a redefinition of scientific methodology. In effect, what was put in question during this period was the status of philosophy of science, as a result of the discovery in Anglophone philosophy of the historicity of science. There is here an important rupture, which today certainly might lead, with the benefit of hindsight, to a discussion of the oft-noted differences between two styles of philosophy of science, French and Anglo-American. The connection of these two traditions in the philosophy of science is perhaps not at first so obvious, nor is it clear how they can engage each other in argument. Kuhn's *Structure of Scientific Revolution* was published in the moribund *Encyclopedia of Unified Science*. Yet the return to French philosophy accomplished by Quine during the 50s and subsequently in a different way by Kuhn, Lakatos, and Feyerabend, was not surprising. This return coincided with, or justified, a new anti-positivism, and a questioning of the “dogmas of empiricism”, Carnap's as well as Popper's, both verificationism and falsificationism.

As Hacking has noted, there is considerable agreement, despite their various differences, between the Logical Empiricists and Popper.

Popper and Carnap assume that natural science is our best example of rational thought [...]. Both think there is a pretty sharp distinction between *observation* and *theory* [...]. Both agreed that there is a fundamental difference between the *context of justification* and the *context of discovery* [...]; the philosophies of Carnap and Popper are timeless: outside time, outside history.⁴⁶

In order to arrive at Kuhn's philosophy, one need only reject these conclusions one by one: Kuhn's claims (like Feyerabend's, though there are differences), even if they are not always framed in this way, bear exactly on this common body of beliefs shared by Carnap and Popper: “whenever we find two philosophers who line up exactly opposite on a series of half a dozen points, we know that in fact they agree about almost everything. They share an image of science”.⁴⁷ It is this image that is drawn into question beginning in the 1960s, precisely with instruments of thought inherited from French philosophy of science – notably that of Duhem and Meyerson. This explains the considerable interest we find in their works during this period.⁴⁸ It is in Duhem, as we have seen, that we find the first formulation of the dependence of experience on theory, whose immediate consequence, recognized by Quine as well as Kuhn, Feyerabend, and Lakatos, is that there is no demarcation between statements of observation and of theory. This dependence of experience on its conceptual context does away with the myths of refutation and crucial experiment. Lakatos takes up this point in “Falsification and the methodology of scientific research programmes”.⁴⁹

Meyerson earlier evokes in *Identity and Reality* “the close dependence of experiments upon scientific theories”.⁵⁰ And later in the *Cheminement de la pensée* Meyerson wrote:

However much one tries to stick to facts, no matter how much effort one makes to exclude every hypothesis, ontology cannot be excluded from physics [...]. Duhem has indisputably shown that an experiment in physics is not just the observation of a phenomenon, but is rather the theoretical interpretation of this phenomenon and that “the statement of the result of an experiment implies, in general, an act of faith in a whole group of theories.”⁵¹

And here it must be recalled that these theories have as their aim precisely to find out the being of things, their essence, and finally to explain the behaviour of objects in terms of this essence, the disposition of particles, of molecules and atoms in bodies, of electrons in the atom.⁵²

It may seem strange at first to see Meyerson appeal to Duhem for purposes contrary to Duhem's own, at least until we look closer: ontology, for Meyerson, is not to be found anywhere except in science ("in terms of this essence, the disposition of particles"). And science cannot perform this work through laws alone, but only by presupposing certain objects (posits, as Quine would say), such as "pure silver [...] the mathematical lever, the ideal gas, or the perfect crystal... abstractions created by the theory."⁵³ For "we only attain laws by violating nature, by isolating more or less artificially a phenomenon from the whole, by checking those influences that would have *falsified* the observation. Thus the law cannot directly express reality."⁵⁴ Meyerson continues his "ontological" interpretation of Duhem in chapter XI of *Identity and Reality*:

Duhem establishes, with great exactness, that only the theoretical interpretation to which phenomena are subjected by the physicist makes possible the use of instruments. He concludes that between phenomena really observed and the result of an experiment formulated by a physicist a very complex intellectual elaboration intervenes.⁵⁵

This argument was reprised by Kuhn and Feyerabend, but Meyerson adds:

These deductions indicate [...] to what a degree the physicist is attached to the concept of *thing* [...]. And it is easy to see why it is impossible to state [an] experiment without speaking of [an] hypothesis. This is because the experiment has to do with something created by this latter; and, of course, the statement when formulated will imply an act of faith in a theory, for it will have to do with the object the essence of which is the basis of the hypothesis in question.⁵⁶

Science creates objects, and it is in *positing* the existence of its objects⁵⁷ (an expression of Meyerson's that Quine takes over) that it achieves its explications. Clearly Meyerson's concept of explication overlaps with the explication rejected by Duhem. More precisely, Meyerson affirms, through his concept of explication, that it was in vain that Duhem tried to exclude all metaphysics from science: metaphysics, or ontology, is natural, immanent in scientific activity. It is on the basis of this philosophical principle that the thesis of *Identity and Reality* is established, namely the constant role of the principle of identity in different stages in the history of science (through principles such as those of *conservation*). The scientist is engaged in ontology, in "pressing his thought into the ontological mold, in giving to it the form of an hypothesis about the reality of things".⁵⁸ He does so without knowing it, like the ordinary man, "comme il respire" (to take up a lovely formulation Meyerson used in *Explication in the Sciences* and quoted by Koyré in his article on Meyerson.⁵⁹ In *Explication in the Sciences* Meyerson wrote that "scientists, as soon as they bring atoms and ether into play, implicitly reason as if these were not concepts, but real things"⁶⁰ In sum, the scientist has a natural tendency to engage in ontology. Meyerson speaks of "the tendency to create fictitious entities for the purpose of explanation", which is "so strongly rooted in us that it was necessary to put us on guard against it by a special declaration [...] the famous 'Ockham's razor'".⁶¹ It is

in this light that we can interpret the central role that the concept of conservation has in Meyerson's work.

Any statement of conservation tends to give rise to an explanatory theory. That is why when confronted with anything that is said to be conserved and which is at first, of course, only a scientific abstraction [...] we feel a sort of irresistible need to hypostasize it ontologically, to transform it into a *being*.⁶²

This tendency to posit theoretical objects explains the origin and development of notions such as that of movement or inertia,⁶³ or simply the transformations of the notion of physical object. In a quite different style, one can recall Quine's Voltairean aphorism: "physical objects, if they did not exist, would [...] have had to be invented".⁶⁴

It is clear from this point of view why Meyerson, as well as Duhem, would be cited by anti-positivism in its opposition to the traditional philosophy of science derived from logical empiricism. We sometimes forget that Meyerson read, at the end of his life, not only the work of Moritz Schlick (at Einstein's urging, he read Schlick's writings on relativity⁶⁵), but also the work of the Vienna Circle and the early Wittgenstein, to which he dedicated a note in *Le Cheminement de la pensée*.⁶⁶ Meyerson was well aware of the proximity of the Vienna Circle (initially called the Ernst Mach Society) to Mach, and therefore to Comte.⁶⁷ In the note, Meyerson expresses surprise that Comte is hardly cited by the Circle. Meyerson is one of the first serious critics of the Viennese tradition, whose theses "on many essential points disagree completely with those [he, Meyerson] presented." It is not at all therefore an accident that Meyerson's critique of positivism reappears later in the post-Popperian critique of neopositivism. What is at stake however, namely realism, is more than just the rejection of positivism, as Meyerson shows in the *Cheminement* by his remarks on Eddington (whose ideas he compares to the realism of Sommerfeld), as well as his lucid critique of both the operationalism of Bridgman and the pragmatism of Dewey.⁶⁸

Meyerson takes up, as we have seen, certain themes from Duhem, but he also connects Duhem with Comte and Mach because of his "phenomenalism": "discussions in physics make no sense if one tries to abandon the assumption that objects exist independent of sensation. The affirmation of the existence of a reality, that never changes".⁶⁹ Duhem admired Mach's *Mechanics*, but Meyerson engaged in a radical critique of Mach in all his works. The most remarkable form of this rejection appears in Meyerson's discussion of Einstein in *The Relativistic Deduction* (1924). The special theory of relativity (1905)⁷⁰ was frequently imagined to be an illustration of the theses of positivism, and later of logical empiricism.⁷¹ The whole aim of *The Relativistic Deduction*, as the title indicates, is to present, contrary to positivism, a deductive, explicative system that posits the existence of an independent reality. Distinguishing clearly between relativity and relativism, Meyerson notes (citing Kneser):

The principle of relativity is, as a matter of fact, the principle of the non-relativity of the real; it demands that the reality implied by the observed phenomena of nature remain immutable with respect to possible modifications of viewpoint and system of measurement, that it be, according to the current expression, invariant with respect to the Lorentz transformations.⁷²

Meyerson chose Einstein to confirm his thesis in *Identity and Reality*. He saw, in the theory of 1905, and later in the general theory of relativity of 1915, the use of a principle of *identity*, which one might also call in this context a principle of invariance. For Meyerson, relativity theory, which proposes the invariance of laws under the Lorentz transformations, just as Galileo had proposed the invariance of laws under change of point of view (for example, that of a sailor on land and of a sailor on a boat in motion), is much more realistic than pre-Einsteinian theories.⁷³ As Sommerfeld, who inspired Meyerson, said: the aim of theories of relativity is to find what is *not* relative. Relativity became for Meyerson the very model of a theory that is explicative and ontological.

This point is yet more interesting when we recall that Einstein himself was won over by Meyerson's interpretation. Until the early 1920s, Einstein presented himself (in accordance with numerous interpretations) as a disciple of Mach, even associating his doctrine, at one time, with what he called *Mach's principle*. It is remarkable that Einstein's turn against Mach, well described by Gerald Holton in his essay "Mach, Einstein, and the Search for Reality"⁷⁴ occurred at the same time as his first contact with Meyerson. They met in 1922, when Einstein was invited to a meeting of the Société Française de Philosophie, and their discussions continued through the publication of *The Relativistic Deduction* and a review of it by Einstein in the *Revue Philosophique*. Einstein's turn against Mach was apparently strengthened by his reading *The Relativistic Deduction*. During the meeting of the Société, Meyerson presented his critique of Mach, and he made it clear that, on his view, "between Mach's ideas and Einstein's theory there seems to be no truly intimate or necessary connection. One can certainly be an adherent of relativity [theory] while being convinced that no science is possible that does not posit, in the first instance, an object persisting outside of consciousness, and that, as a consequence, science cannot avoid the task of making clear how it conceives this object, through the modifications that the progress of our knowledge imposes on this image. Indeed it seems to me that Einstein's attitude confirms this point of view".⁷⁵ On the same occasion Einstein objected to Mach as a philosopher:

Mach's system studies relations that exist among the data of experience; the totality of these relations is, for Mach, science. But this is a bad point of view: in sum, what Mach did is a catalogue and not a system. Mach was as awful at philosophy as he was good at mechanics.⁷⁶

This surprising claim shows clearly that there is a connection between Meyerson's views and Einstein's move toward realism. If this connection is not one of cause and effect, it is at least one of convergence, Einstein finding in Meyerson terms and arguments appropriate for his rejection of Mach's views.

Meyerson clearly influenced Einstein's philosophical development. One can particularly trace this influence in his correspondence with his friend Michel Besso, a convinced Machian. It is in a 1917 letter to Besso that Einstein first presents his doubts about Mach, in a discussion of a manuscript by Friedrich Adler (physicist, Austrian politician, and translator of Duhem). Regarding Mach's philosophy, Einstein wrote: "It cannot give birth to anything living, it can only exterminate harmful vermin".⁷⁷ In a much later letter to Besso (1948), he elaborates on this point, in a tone remarkably reminiscent of Meyerson:

It is interesting, by the way, that Mach rejected the special relativity theory passionately [...]. The theory was, for him, inadmissibly speculative. He did not know that this speculative character belongs also to Newton's mechanics, and to every theory which thought is capable of. There exists only a gradual difference between theories, insofar as the chains of thought from fundamental concepts to empirically verifiable conclusions are of different lengths and complications.⁷⁸

Meanwhile, in a letter to Schlick Einstein seems to change course and follow Meyerson in opposing neopositivism. He reproaches Schlick:

In general, your presentation does not correspond to my style of thinking, in as much as I find your overall orientation so to speak too positivist [...]. I put it to you squarely: physics is an *attempt at the conceptual construction of a model of the real world and its nomological structure* [...]. In sum, I object to the failure to clearly separate *the reality of experience and the reality of being*.

One cannot help but notice the similarity between the realist positions defended by Einstein and by Meyerson. This similarity is confirmed if we keep in mind not only relativity theory, but also Meyerson's position on the Copenhagen interpretation of quantum physics. Meyerson's reservations about the latter were, like Einstein's, on ontological grounds: in *Réel et déterminisme dans la physique quantique* (1933), he affirms that Bohr and Heisenberg could not have done otherwise than to posit an independent reality: "The quantum physicist, in as much as he is a physicist, certainly thinks as a realist, cannot think otherwise than as a realist".⁷⁹ "Quantum physics, like any other physics, presupposes a real outside of me".⁸⁰ Phenomenalist interpretations are, for Meyerson, an admission of failure, in fact "a sort of homage paid to the idea of a real explication": "If the least possibility offers itself, we see researchers come back to a concrete image, realizable in thought, a *Weltbild*".⁸¹ Meyerson's thought can thus help us understand the particular brand of realism gradually adopted by Einstein, which one might call, adopting an expression of Putnam's, *natural realism*, or following Arthur Fine, a *natural ontological attitude*. This realism affirms the necessity, presented by Meyerson in *Identity and Reality* and developed in the *Cheminement de la pensée*, of an ontology inherent in science:

This [ontological] aspiration is entirely supported by science, which, in this respect as in many others, is *nothing but a particular form of philosophy*, but a philosophy necessarily realist, incapable of divesting itself of an ontology.⁸²

Einstein, who, in his review of *The Relativistic Deduction*, recognized only that "All science is founded on a realistic philosophical system", went further (and here again we see the influence of Meyerson) in a letter to Schrödinger (1935): "The real problem is that *physics is a kind of metaphysics*; physics describes 'reality'. But we do not know what 'reality' is. *We know it only through physical description*".⁸³

The problem here posed goes beyond the context of Einstein's philosophy, as the debates over realism in the late 20th century show: in affirming the connection between science and reality, we hesitate between an immanent ontology and a robust realism that claims this "posited" reality as the only reality. This contradiction, which structures the whole of Quine's work, was formulated by Meyerson in *The Relativistic Deduction*. Meyerson recognized that "Although the Einsteinian physicist, like all physicists, is basically a realist, the very suc-

cess of his deduction leads him to a structure that is just as basically idealistic".⁸⁴ Quine, in 1950, taking up in a naturalist vein the phrase "L'ontologie fait corps avec la science elle-même et ne peut être séparée" (In French in Quine's text : "Ontology is part of the body of science itself and cannot be separated from it")⁸⁵ and inscribing it in the context of a relativized ontology (it is science that tells us *what exists*) could not but end up with a radicalized form of the duality of realism described by Meyerson. The realist side of Quine, constantly reiterated in his work, applies only locally. The conceptual scheme, "the whole scientific system, ontology and all, is a conceptual bridge made by us". Realism is "robust" because it is immanent to our language and to our understanding of science. The content that we give to the word "reality" is produced by our scientific discourse, and integrated into an "immanent epistemology". This is the limit of Meyerson's influence, and the full radicalism of Quine, who owes to Meyerson even his conception of naturalism.

Meyerson is far from thinking, contrary to Duhem or to certain interpretations of Duhem, that science and its ontology are exempt from testing by experiment. "No one will dream of developing a scientific theory without showing to what extent it is confirmed by experience".⁸⁶ It is in the face of recalcitrant experience that changes of theory occur. We find this aspect of Meyerson's philosophy again in Kuhn, who is not really the idealist one often supposes from a cursory reading of *The Structure of Scientific Revolutions*. Kuhn affirms that a change of theory does not occur except when there is a general recognition of difficulties and failures of the current theory, what he calls anomalies.⁸⁷ In the essay "Hegel, Hamilton, Hamelin, et le Concept de Cause",⁸⁸ Meyerson takes up from this point of view the adage "a theory is no good unless one can show that it is false". He writes:

This is evident for a scientific theory that accommodates itself to no matter what observations and experiments is a theory that is so flexible and inclusive as to be decrepit; it is useless even from the point of view of the simple prediction of facts, and does not persist even for a moment unless there is no other to put in its place.

Thus Meyerson anticipates, here again, the anti-positivist reactions of Kuhn and Feyerabend. Meyerson, like Duhem, affirms that an isolated experiment cannot suffice to refute a theory; but this is because, for him, a theoretical change is *also* an ontological change. This leads him to formulate a conception of the history of science close to that of Feyerabend, for whom scientific changes do not take place in the absence of an "alternative" theory, and to that of Kuhn, for whom a paradigm is not rejected unless it is in a lamentable state.

Thus a physical theory, as is easily shown by an examination of the whole history of the sciences, does not disappear unless it is succeeded by a new theory; the scientific reality that dies is of necessity born again in a new reality.⁸⁹

Far from being the simplistic continuist one might suppose, Meyerson proposes a reading of the history of science that is in fact the true precursor of the 1960s.

3 Toward Anthropology

Meyerson was without doubt one of the first, with Duhem, to see the true nature of scientific change as it was later explored by Kuhn. In a paragraph of the *Cheminement de la pensée*, titled “Les Révolutions dans les sciences physiques” he suggests that the history of science, as it is usually presented, leaves out the resistance that always opposes new ideas, or presents it “in a way that only the innovator himself would find justified, his opponents appearing as men of ill-will, or of mediocre intelligence, incapable of grasping the clearest evidence”. According to Meyerson, we must re-examine history and recognize the resistance of normal science to change:

And if one takes the trouble simply to examine, without preconceived notions, the polemics of this great period, one quickly sees where the resistances come from and that none of them is without possible justification.⁹⁰

Elsewhere he elaborates on this point, criticizing the usual reading of the chemical revolution:

The arguments of the phlogiston chemists were in no way absurd, nor were they unscientific (contrary to what men insufficiently informed in the documents of the history of science have often maintained) [...]. Lavoisier violated the most essential rules of chemical argument as they were firmly established at that time.⁹¹

Here we find exactly the formulation of the proprieties of normal science as Kuhn defined it in *Structure*. Perhaps it was through Koyré that this specific mode of interpreting historical scientific texts was transmitted, as many of the essays collected by Kuhn in *The Essential Tension* suggest.

This is precisely the problem of incommensurability. From the very beginning of *The Structure of Scientific Revolutions*, the theme of “incommensurable ways of viewing the world” is raised by the problem of historians’ access to “how things were before”, and of the scientificity of past science.

Historians confront growing difficulties in distinguishing the “scientific” component of past observation and belief from what their predecessors had readily labeled “error” and “superstition”.⁹²

It is a matter of considering past theories with the attention that Meyerson and Koyré advocated, not as receptacles of error, but as part of science. What Kuhn and Feyerabend most object to in philosophy of science before them is not so much its rationalism as its conception of past theories as errors, and of history as a succession of refutations and corrections, even provisional ones. From this point of view we can understand better the meaning of a remark of Hacking, for whom the center of the philosophical revolution introduced by Kuhn is “a different relation of science to its past”. This is not a matter of a formal respect for the past, a kind of principle of charity adapted to the history of science; rather it is a matter of a different relation to experience:

If these out-of-date beliefs are to be called myths, then myths can be produced by the same sorts of methods and held for the same sorts of reasons that now lead to scientific knowledge. If, on the other hand, they are to be called science, then science has included bodies of belief quite incompatible with the ones we hold today. Given these alternatives, the his-

torian must choose the latter. Out-of-date theories are not in principle unscientific because they have been discarded.⁹³

The thesis of incommensurability, for Kuhn, applies to paradigms of both intelligibility and rationality: each scientific revolution displaced “the standards by which the profession determined what should count as an admissible problem or as a legitimate problem-solution”.⁹⁴ But paradoxically Kuhn sees the need for, as Koyré had already suggested, a principle of universal intelligibility.

What has made the assumption of universal translatability so nearly inescapable is, I believe, its deceptive similarity to a quite different one, in this case an assumption that I share: anything which can be said in one language can, with imagination and effort, be *understood* by a speaker of another. What is prerequisite to such understanding, however, is not translation but language learning. Quine’s radical translator is, in fact, a language learner.⁹⁵

Returning again to Quine, we see that what is at stake here is the anthropological dimension of the question of incommensurability. The principle of identity, elaborated in Meyerson’s early works, brings anthropology and philosophy of science together in a particularly fruitful way, just as it did in interesting discussions between Meyerson and Lévy-Bruhl. A whole chapter of the *Cheminement de la pensée* is dedicated to the connection between “The physicist and primitive man”.⁹⁶ In light of Lévy-Bruhl’s work on participation and his interpretation of pre-logical mentality, Meyerson argues that the scientist of the past, like the primitive, “did not depart for all that from the general stamp of our intellect”.⁹⁷ When we attribute to the primitive (or the past scientist) a mode of thought different from ours, we refuse to see that he reasons as we do: “The primitive judged wrongly, but he nonetheless thought as we habitually do, and we cannot pretend that he was illogical without affirming at the same time that our own way of thinking is too”.⁹⁸

If we follow Lévy-Bruhl, then (despite many interpretations to the contrary) we must attribute to the primitive a form of common rationality. The question of logic is to be posed, not at the level of individual psychology, but at the level of a comparative study of diverse types of collective mentality. It is this comparativism that determines Lévy-Bruhl’s method, and it is no way relativist: it is more a matter of showing the difficulty of defining logic once one gives up trying to ground or define it in terms of a single type of human mind or a transcendent rationality:

I do not assert (today less than ever) that there exists a mentality peculiar to “primitive peoples”. There is, in their mentality, a large part which they have in common with us. Equally there is in the mentality of our societies a part (larger or smaller according to the general conditions beliefs, institutions, social classes, etc.) which is common to it and to that of primitive peoples.⁹⁹

Ethnography does not aim to establish either insurmountable differences in thought, or the psychological unity of the human species. It aims to bring to light, by the affirmation of what is shared between the primitive and the non-primitive, an *immanent plurality* in thought. And this is precisely the way, according to Meyerson, and later, Kuhn, that the history of science should proceed. The connection that Meyerson’s principle of identity creates between anthropology and philosophy of science makes *identity* a condition on the discovery of conceptual *diversity*. In an essay on the interest of Lévy-Bruhl’s theory of participation for the history of science Héléne Metzger has discussed

this similarity between the work of the ethnologist and that of the historian of science, in their study of universal schemas that *make it possible to cognize differences*.¹⁰⁰

The appeal to anthropology also allows us to de-dramatize relativism, and especially to defuse the Davidsonian critique of “conceptual schemes” and paradigms as sources of relativism. In his “Reflections on My Critics”¹⁰¹ Kuhn replies to a similar objection from Popper, who criticizes the “dogma [...] [that] different frameworks are like mutually untranslatable languages”.¹⁰² Kuhn recognizes that we can only examine the paradigms of the past from within our own, but he argues that this does not keep us from examining them. This is precisely the work of the historian of science. To do the history of science is to learn how to translate historical languages into our terms. We translate “ancient theories in modern terms”, and we learn, for example, to read historical documents differently. “Part of learning a language or a theory”, Kuhn wrote, “is learning to describe the world within which the language functions” and to “acquire the knowledge of nature that is built into the language”.¹⁰³ The new task given to the historian of science resembles, for Kuhn, not so much translation as the learning of a foreign language or culture, which is perfectly accessible provided we perceive the distance between a distant paradigm and ours and learn where they differ. With time, we can learn the language of the other culture in this way, which does not mean interpreting or conceptualizing it in our language, but rather learning to predict the reactions of the other, and making his strangeness familiar – “something that the historian regularly learns to do (or should) when dealing with older scientific theories”.¹⁰⁴ This conception of incommensurability has nothing in common with extreme relativism: it recognizes a new task for philosophy of science, a descriptive one, in showing that the only way to describe the experience of the other is to take the measure of his distance from us.

The work of the historian of science thus turns out to be similar to that of Quine’s linguist undertaking radical translation: it is a matter of reading a foreign language in order to give it meaning, to integrate it into our language. The possibility of translation is the basis not only of the history of science but also of the growth of knowledge. There is an indeterminacy to translation, but it is this indeterminacy that makes the growth of science possible. This point was also made by Koyré, in his beautiful, and very critical, review of Louis Rougier’s book, *La Scolastique et le Thomisme*.¹⁰⁵

Nothing is more variable than the collections of “truths” admitted and believed at different times by different social groups. Again, nothing is more variable than mental attitudes, both individual and social [...]. It is obvious that a primitive who believes in magical causes [...] and a physicist who studies the laws of motion have quite different mental attitudes [...]. But inside their mentality and their beliefs they think [...] in the same way. Despite the *material* differences there is a *formal* identity of thought. This is not, I believe, an *a priori* claim. The profound analyses of Lévy-Bruhl on the one hand, and of Meyerson on the other, have, I believe, firmly demonstrated this formal identity of the categories of thought.¹⁰⁶

If we translate correctly the propositions “written by those who preceded us”, that is to say, “if we are willing to seek out the theories that give these propositions their true sense”, then we can “translate them into the language of the theories

accepted today” and see the truth in them. Meyerson takes up the same theme in an even more optimistic form, and reverses Duhem’s conclusion:

The science of the past is every bit as useful as that of today for the study of these processes. One might even say more useful. For by the very fact that this science is *outdated*, that we no longer believe in it, we are able to observe it more impartially. Indeed, however hard we try, we cannot attain such impartiality toward the science of today. The latter, its methods and its results, are among the most essential components of our intellectuality.¹⁰⁷

Meyerson proposes, as a criterion of translation, the “identity” of the human mind.

It is here that the history of science is in danger of making us feel awkward, since it shows a thought process whose course follows the same principles as ours does, yet the conclusions it arrives at are so different from those we are used to.¹⁰⁸

The parallel between history of science and anthropology has proven itself especially fruitful, for Koyré and others from Quine and Kuhn to Foucault, and it seems to me that the whole approach implicit in these remarks, one that connects anthropology and history of science, defines a theme specific to contemporary philosophy, and maybe one of its central inspirations.¹⁰⁹

Endnotes

- 1 “Entretien de Christian Delacampagne avec Thomas S. Kuhn”, *Le Monde*, 5–6 Feb., 1995.
- 2 Until a recent revival, thanks to a group of French researchers : see e.g. Brenner, 1990, 2003, Bitbol and Gayon, 2006.
- 3 Willard Van Orman Quine, “Two Dogmas of Empiricism”, in Quine, 1961, p. 41.
- 4 Duhem, 1906; transl. pp. 183–190.
- 5 *Ibid.*, transl. pp. 190–208.
- 6 *Ibid.*, transl. p. 159.
- 7 *Ibid.*, transl. p. 163.
- 8 Meyerson, 1908; transl. p. 31.
- 9 Quine, 1982, p. 2.
- 10 Meyerson, 1931, p. 125.
- 11 Duhem, 1906; transl. p. 184.
- 12 Quine, 1982, p. 2.
- 13 Duhem, 1906; transl. p. 187.
- 14 Quine, 1961, p. 42.
- 15 Duhem, 1906; transl. p. 216.
- 16 Quine, 1961, p. 42.
- 17 Quine, 1982, p. 2.
- 18 *Ibid.*
- 19 *Ibid.*
- 20 See Quine 1961, p. 79.
- 21 *Ibid.*
- 22 See *Ibid.*, p. 46.
- 23 Quine, 1960, p. 4.
- 24 Duhem, 1906; transl. pp. 187–188.
- 25 *Ibid.*, transl. p. 188.
- 26 *Ibid.*, transl. p. 26.

- 27 Quine, 1961, p. 19.
- 28 Ibid., p. 44.
- 29 Ibid., p. 44.
- 30 Carnap, 1934; transl. p. 318.
- 31 Quine quotes this sentence from *Identité et réalité* (1961, note 20, p. 45).
- 32 Quine, 1961, p. 103.
- 33 Quine, 1969, p. 93. Emphasis mine.
- 34 Quine, 1960, p. 275.
- 35 See Laugier 1992.
- 36 Meyerson, 1931, p. 125. All translations from this text J. E.
- 37 Meyerson, 1908, p. 385.
- 38 Ibid., p. 435.
- 39 Quine, 1981, p. 22.
- 40 Quine, 1969, p. 98.
- 41 Quine, 1981, p. 21.
- 42 Ibid., p. 67.
- 43 Quine, 1960, p. 24. See also 1961, p. 79.
- 44 Quine, 1995, p. 260.
- 45 Ibid.
- 46 Hacking, 1983, pp. 5–6.
- 47 Ibid.
- 48 *Identité et réalité* was first translated into English in 1930 (Folcroft); *La déduction Relativiste* (with Einstein's review as an appendix) in 1985 (Reidel); *L'explication dans les sciences* in 1991 (Kluwer). Duhem's trans-Atlantic success, though it came later, was even greater: A translation of *La théorie physique (The Aim and Structure of Physical Theory)* appeared in 1962 (Atheneum), *L'évolution de la mécanique* in 1980 (Kluwer), *Sôzein ta phainomena (To Save the Phenomena)* in 1985 (University of Chicago, reprinted in paperback), excerpts from *Système du monde* (as *Medieval Cosmology*) in 1987 (University of Chicago), *Les origines de la statique* (Kluwer, 1991) and even *La science allemande (German science, Open court, 1991)*. During this time only Meyerson's *L'explication dans les sciences* (Corpus) and Duhem's *La théorie physique* (Vrin), *L'évolution de la mécanique* (Vrin), and *Le Système du monde* (Hermann) have been in print in their original language. An abridged French version of *Le Système du monde* was edited by Anastasios Brenner under the title *L'aube du savoir* (Hermann, 1997).
- 49 Lakatos, 1995.
- 50 Meyerson, 1908, pp. 391–392. This passage probably influenced Quine, and may have led him to later read Duhem.
- 51 Cf. Duhem 1906, transl. p. 183.
- 52 Meyerson, 1931, p. 125.
- 53 Meyerson, 1908, p. 31.
- 54 Ibid.
- 55 Ibid., p. 368. Translation modified.
- 56 Ibid.
- 57 Ibid.
- 58 Ibid., p. 387.
- 59 Koyré, 1931, p.217.
- 60 Meyerson, 1921, p. 85.
- 61 Ibid., p. 61.
- 62 Ibid., p. 120.
- 63 See Meyerson, 1925.
- 64 Quine, 1982, p. 1.
- 65 See Schlick, "The Philosophical Significance of Relativity" and "Space and Time in Contemporary Physics" in *Philosophical Papers*, vol. 1.
- 66 Meyerson, 1931, pp. 787–790.

- 67 Ibid., p. 115.
 68 Ibid., p. 25. See also pp. 796–800.
 69 Ibid., p. 118.
 70 See Einstein, 1987, *Collected Papers* vol. 2 for discussions.
 71 See the works of Petzoldt, Schlick, and P. Frank. See also Meyerson, 1925.
 72 Meyerson, 1925, p. 50.
 73 Ibid., p. 55.
 74 Holton, 1968, pp. 637–673.
 75 Einstein, 1922, p. 111.
 76 Ibid., pp. 111–112.
 77 Quoted in Holton, 1968, p. 657.
 78 Quoted in Ibid., pp. 648–649.
 79 Meyerson, 1933, p. 48.
 80 Meyerson, 1931, p. 118.
 81 Ibid., p. 49.
 82 Ibid., p. 117
 83 Quoted in Arthur Fine, 1986, p. 125. Emphasis mine.
 84 Meyerson, 1925, p. 99.
 85 Meyerson, 1908, p. 7.
 86 Meyerson, 1921, p. 514.
 87 Kuhn, 1977, pp. 202–211.
 88 *Revue philosophique* 96 (1923), pp. 33–55.
 89 Meyerson, 1931, p. 589.
 90 Ibid., pp. 547–548.
 91 Ibid., p. 483.
 92 Kuhn, 1962, p. 2.
 93 Ibid., p. 2–3.
 94 Ibid., p. 6.
 95 Kuhn, 2000, p. 61.
 96 Meyerson, 1931, pp. 49–88.
 97 Ibid., p. 83.
 98 Meyerson, 1931, p. 83.
 99 Lucien Lévy-Bruhl, 1975, pp. 125–126. Translation modified by J. Elliott
 100 Metzger, 1930, pp. 15–24.
 101 Lakatos and Musgrave, 1970, p. 267.
 102 Ibid., p. 56.
 103 Ibid., p. 270.
 104 Ibid., p.277.
 105 Paris, Gauthier Villars, 1925.
 106 *Revue philosophique* 99 (1926), p. 466.
 107 Meyerson, 1921, p. 528.
 108 Meyerson, 1931, p. 85.
 109 Many thanks to Jay Elliott for his work on the English version of this text, and to Michel Bitbol, Alain Boyer, Christian Bonnet, Anastasios Brenner, Frédéric Fruteau de Laclos, Isabelle Peschard, Elie Zahar for fruitful discussions of Meyerson’s work, and of various versions of this paper.

Bibliography

- Bitbol, Michel et Gayon, Jean (2006), *L’Épistémologie française, 1830–1970*, Paris, PUF. 2006.
 Brenner, Anastasios (1990), *Duhem : science, réalité et apparence. La relation entre philosophie et histoire dans l’œuvre de Pierre Duhem*, Paris, Vrin, »Mathesis«.

- Brenner, Anastasios (2003), *les origines françaises de la philosophie des sciences*, Paris, PUF.
- Carnap Rudolf (1934), *Logische Syntax der Sprache; The Logical Syntax of Language*, transl. Amethe Smeaton, London, Kegan Paul, 1937.
- Duhem Pierre (1903), *L'Évolution de la mécanique*, Paris, Vrin, 1992; *The Evolution of Mechanics*, transl. M. Cole, Alphen aan den Rijn, Sijthoff and Noordhoff, 1980.
- Duhem Pierre (1906), *La Théorie physique, son objet et sa structure*, Paris, Vrin, 1981; *The Aim and Structure of Physical Theory*, transl. P. Wiener, Princeton, N.J., Princeton University Press, 1962.
- Duhem Pierre (1905–1906), *Les origines de la statique*, 2 vol., Paris, Hermann. *The Origins of Statics*, transl. G.F. Leneaux et al., Kluwer, 1981.
- Duhem Pierre (1908), *Sauver les phénomènes. Essai sur la notion de théorie physique de Platon à Galilée*, Paris Vrin, 1982. *To Save the Phenomena*, transl. E. Doland and C. Maschler, Chicago, Chicago University Press, 1969.
- Duhem Pierre (1913–1959), *Le système du monde. Histoire des doctrines cosmologiques de Platon à Copernic*, 10 vol., Paris, Hermann; *Medieval Cosmology*, partial transl. Roger Ariew, Chicago, Chicago University Press, 1985.
- Duhem Pierre (1915), *La science allemande*, Paris, Hermann; *German Science*, transl. J. Lyon, La Salle, Open Court, 1991.
- Duhem Pierre (1997), *L'aube du savoir: épitomé du Système du monde*, A. Brenner, (ed.), Paris, Hermann.
- Einstein Albert (1987), *Collected Papers*, vol. 2, Princeton, N.J., Princeton University Press.
- Fine Arthur (1986), *The Shaky Game: Einstein, Realism, and the Quantum Theory*, 2nd ed., Chicago, University of Chicago Press, 1996.
- Hacking Ian (1983), *Representing and Intervening*, Cambridge, Cambridge University Press.
- Holton Gerald (1968), "Mach, Einstein and the Search for Reality", *Daedalus*, 97, pp. 637–673.
- Koyré Alexandre (1931), "Die Philosophie Emile Meyerson", *Deutsch-Französische Rundschau*, 4, pp. 197–217.
- Koyré Alexandre (1926), "Compte rendu de *La Scolastique et le thomisme*" par L. Rougier, Gauthier Villars, 1925, *Revue philosophique*, 1926, pp. 462–468.
- Kuhn Thomas (1962), *The Structure of Scientific Revolutions*, Chicago, University of Chicago Press.
- Kuhn Thomas (1977), *The Essential Tension*, Chicago, University of Chicago Press.
- Kuhn Thomas (2000), *The Road since Structure*, James Conant and John Haugeland (eds.), Chicago, University of Chicago Press.
- Lakatos Imre (1995), "The methodology of scientific research programmes", *Philosophical papers*, volume I, edited by John Worrall and Gregory Currie, Cambridge, Cambridge University Press.
- Lakatos Imre, Alan Musgrave (eds.), (1970), *Criticism and the Growth of Knowledge*, Cambridge, Cambridge University Press.
- Laugier Sandra (1992), *L'anthropologie logique de Quine*, Paris, Vrin.
- Laugier Sandra (2003) (ed.), "Après la Structure – Kuhn, les révolutions scientifiques et l'incommensurabilité", *Archives de Philosophie*.
- Lévy-Bruhl Lucien (1949), *Les Carnets*, Paris, Presses universitaires de France; *The Notebooks on Primitive Mentality*, transl. Peter Rivière, New York, Harper and Row, 1975.
- Quine Willard Van Orman (1960), *Word and Object*, Cambridge, Massachusetts, MIT Press.
- Quine Willard Van Orman (1961), *From a Logical Point of View*, 2nd ed., Cambridge, Massachusetts, Harvard University Press.
- Quine Willard Van Orman (1969), *Ontological Relativity*, New York, Columbia University Press.
- Quine Willard Van Orman (1981), *Theories and Things*, Cambridge, Massachusetts, Harvard University Press.
- Quine Willard Van Orman (1982), *Methods of Logic*, 4th ed., Cambridge, Massachusetts, Harvard University Press.
- Quine Willard Van Orman (1995), "Naturalism; Or, Living Within One's Means", *Dialectica* 49/2, 251–261.

- Metzger Hélène (1930), “La philosophie de Lévy-Bruhl et l’histoire des sciences”, *Archeion*, 12, pp. 15–24.
- Meyerson Émile (1908), *Identité et réalité*, Paris, Vrin, 1951; *Identity and Reality*, London, George Allen and Unwin, 1930.
- Meyerson Émile (1921), *De l’explication dans les sciences*, 2 vols, Paris, Payot; transl. *Explanation in the Sciences*, Dordrecht, Kluwer, 1991.
- Meyerson Émile (1923), “Hegel, Hamilton, Hamelin et le concept de cause”, *Revue philosophique*, 96, pp. 33–55.
- Meyerson Émile (1925), *La Déduction relativiste*, Paris, Payot; transl. *The Relativist Deduction: Epistemological Implications of the Theory of Relativity*, Dordrecht, Reidel, 1985.
- Meyerson Émile (1931), *Le Cheminement de la pensée*, Paris, Alcan.
- Meyerson Émile (1933), *Réel et déterminisme dans la physique quantique*, Paris, Hermann, “Actualités scientifiques et industrielles”.
- Rougier Louis (1925), *La Scolastique et le thomisme*, Paris, Gautier Villars.
- Schlick Moritz (1979), *Philosophical Papers*, Henk L. Mulder and Barbara F.B. Van de Velde-Schlick (eds.), transl. Peter Heath, 2 vol., Dordrecht, Reidel.

Part III
Physical and Chemical Sciences

Philosophy and 20th Century Physics

Gilles Cohen-Tannoudji

1 Introduction

The celebration in 2005 of the centennial of the “miraculous year” during which Einstein produced his articles on the energy quanta, on the Brownian motion and on restricted relativity has provided an opportunity to draw up a comprehensive assessment of the contribution of 20th century physics to human knowledge. One must recognize that this contribution is impressive. Contemporary physics has made available what is known as the *standard model*, namely, a set of *effective theories* that, with the help of a finite set of adjustable parameters, lead to an acceptable agreement with all experimental or observational data on the microscopic structure of matter and on the evolution of the universe.

The study of the microscopic structure of matter is the objective of the physics of elementary particles and fundamental interactions. This part of physics is the heir of the atomistic conception of ancient Greek philosophers, according to which all the various forms of matter are determined by the combinatorial arrangements of huge numbers of infinitesimal, irreducible constituents that exist in a small number of different species, and, as such, it has far reaching philosophical implications. In this domain, the standard model consists, on the one hand, of quantum Chromo Dynamics (QCD), the theory of the strong interactions of *quarks* and *gluons*, and, on the other hand, of the *electroweak theory* of the electromagnetic and weak interactions of *quarks*, *leptons*, *intermediate* and *Higgs bosons*. The theoretical framework of this part of the standard model is the *quantum theory of fields*¹ that realizes the merging of quantum physics and restricted relativity.

The study of the universe as a whole is the objective of cosmology, a domain that, until recently, belonged rather to philosophy than to science. It is not the least merit of 20th century physics to have provided this domain with a scientific basis through Einstein’s theory of general relativity.² This theoretical framework has

G. Cohen-Tannoudji (✉)
Laboratoire de recherche sur les
sciences de la matière (LARSIM)
CEA Saclay
91191 Gif-sur-Yvette cedex, France

made it possible to put together the observational data in a cosmological standard model, the so-called *big bang* model.

The standard models of particle physics and of cosmology both involve a time–energy relation: in particle physics that belongs to quantum physics, this relation is a consequence of the Heisenberg inequalities stating that the product of indeterminacies on the measurement of time and space variables and those on the measurement of energy and momentum variables is bound to be larger than the quantum of action equal to Planck’s constant \hbar ; in cosmology, according to the big bang model, the universe is expanding, diluting and cooling after an initial singularity, the big bang, when it was infinitely dense and hot; in its primordial state, the universe is modeled as a homogeneous fluid the temperature of which, namely the average kinetic energy of its constituents, decreases as the inverse of the square root of the time elapsed since the big bang. Due to this circumstance, particle physics and cosmology acquire, through their convergence, a fascinating temporal dimension: exploring the world of the infinitely small with a high energy probe amounts to simulate, in the laboratory, the conditions prevailing in the primordial universe, at a time after the big bang when the temperature corresponded to the energy of the probe. The representation of the universe that the standard models of particle physics and of cosmology are offering us is one of a universe in evolution, in becoming, from a primordial phase when all interactions and particles were unified to the state in which it can now be observed through a long sequence of phase transitions in which interactions differentiate, particles acquire their masses, symmetries are broken, new structures form, new states of matter emerge. In this exploration one has to rely on the methods of statistical physics, a domain in which important philosophical questions arise. In any case, again, physics is getting a foothold in a domain that *par excellence* belongs to philosophy, namely *cosmogony*.³

The objective of the present chapter is to present to a public of philosophers of science the philosophical implications of 20th century physics as a physicist understands them. We shall have to discuss the fundamentals of the theoretical framework of the standard model in connexion with some philosophical issues concerning reality, objectivity, causality, and completeness, arrow of time, reductionism, and determinism. For this discussion we shall rely heavily on the contribution of Einstein not only because he has initiated almost all the developments of 20th century’s physics, but also because his epistemological⁴ writings, including his acute criticism of quantum physics provide very useful guiding lines for those who want to understand the philosophy of contemporary physics. We shall first recall the program of *rational mechanics* whose aim was to comprehend the whole of physical reality in terms of the motion of material objects in space and time, and that developed from the works of Newton to the apogee of the end of the 19th century. We shall then describe the *deep conceptual crisis* this program went through at the beginning of the 20th century, and then explain the very profound *transformations of the conceptual basis* of physics called for by this crisis and validated by the successes of the standard models. It is this validation by the confrontation of theory and experiment that enables us to reach a reliable understanding of the philosophical implications

of modern physics. At the beginning of this chapter I wish to apologize for some technicalities in the following developments that may seem hard to follow for a non-specialist: indeed I believe that the price to pay for this reliable understanding is to be at least aware of the real stakes of the conceptual developments that led to the current achievements.

The problems encountered by the founders of contemporary physics were extremely difficult because of their far-reaching philosophical implications. To solve these problems, physicists could not and did not want to rely on any philosophical system, because the very adherence to a system would have restricted the field of possibilities in the search for a way out of the conceptual crisis they were confronting. This attitude toward philosophical systems, which I shall adopt in the present chapter, is very well-expressed by Einstein in his “Reply to Criticisms”, included in *Albert Einstein: Philosopher-Scientist*:

The reciprocal relationship of epistemology and science is of noteworthy kind. They are dependent upon each other. Epistemology without contact with science becomes an empty scheme. Science without epistemology is — insofar as it is thinkable at all — primitive and muddled. However, no sooner has the epistemologist, who is seeking a clear system, fought his way through to such a system, than he is inclined to interpret the thought-content of science in the sense of his system and to reject whatever does not fit into his system. The scientist, however, cannot afford to carry his striving for epistemological systematic that far. He accepts gratefully the epistemological conceptual analysis; but the external conditions, which are set for him by the facts of experience, do not permit him to let himself be too much restricted in the construction of his conceptual world by the adherence to an epistemological system. He therefore must appear to the systematic epistemologist as a type of unscrupulous opportunist: he appears as realist insofar as he seeks to describe a world independent of the acts of perception; as idealist insofar as he looks upon the concepts and theories as the free inventions of the human spirit (not logically derivable from what is empirically given); as positivist insofar as he considers his concepts and theories justified only to the extent to which they furnish a logical representation of relations among sensory experiences. He may even appear as Platonist or Pythagorean insofar as he considers the viewpoint of logical simplicity as an indispensable and effective tool of his research.⁵

2 The Program of Rational Mechanics

The general program of mechanics known as *rational*, initiated by the works of Galileo and Newton, marks, in the aftermath of the Renaissance, what one can call the birth of modern science. This program consists in trying to reduce the whole of physics to mechanics, i.e. to the study of the motion of material objects in space and time. The two basic concepts of the program of mechanics are the *material point* and the *force*, starting points of the two roads that lead to the current physical concepts of *elementary particle* and *fundamental interaction*. The concept of material point is a sort of asymptotic concept: it corresponds to the simplest material object the motion of which in space and time can be determined according to the program of mechanics. It obviously corresponds to the atomistic intuition of elementary, point-like, structure-less constituents of matter, which implies that eventually the

program of mechanics will converge with the atomistic conception of the world. The concept of force, on the other hand, is somehow the blind spot of the program. In fact, in rational mechanics, forces are supposed to be given, they are not the object of any theoretical derivation, to use a common terminology, “they are put by hands” (“*hypotheses non fingo*” says Newton), they can act instantaneously at a distance. In mathematized mechanics, forces are often taken as deriving from a *potential*. The program of mechanics can then be reduced to the two following reciprocal questions:

- Given a system of material points, and some forces, what motion do these forces induce for the system of material points (provided that the initial conditions are fixed)?
- Given the motion of some material points, what are the forces that have given rise to this motion?

The immense success of the program of mechanics, in particular when it was applied to the motion of planets, is incontestably due to the effectiveness of its mathematical method. Newton is indeed the founder, at the same time as and independently of Leibniz, of what one now calls the differential and integral calculus, which enabled him to develop the mathematical formalism of mechanics. Under the action of the continuators of Newton, like Euler, Lagrange, Hamilton and Jacobi, rational mechanics, developed considerably, and reached, at the end of the 19th century a true apogee. It is interesting to note that in spite of the crisis it went through at the beginning of the 20th century, the ambition of mechanics remains a true guiding principle of research in contemporary theoretical physics.

Intended at the beginning to account for the motion of simple material points, mechanics immediately tackled the description of the most general motions affecting material objects of any kind. After the material point, the simplest object that one can consider is the rigid solid body, the motion of which is split into the translation motion of its centre of mass, and a rotational motion around this centre of mass. Mechanics extends then to the dynamics of fluids, which one decomposes by thought into infinitesimal cells comparable to material points. It thus appears that with the concepts of material point and force, mechanics has vocation to extend to the description of the sum total of all physical phenomena, provided that one carries out the extension of its applicability to phenomena like light, electricity, magnetism or heat.

Such an extension of mechanics obviously required empirical or experimental explorations but also the significant improvements of the formalism of mechanics that one owes to the above-mentioned continuators of the work of Newton. Lagrange thus revolutionizes mechanics by axiomatizing it in what he calls *analytical mechanics*. He unifies mechanics mathematically, by establishing a formal framework making it possible to solve all the problems of mechanics, including statics and dynamics, for solids or fluids. This reformulation of mechanics ascribes a central role to the concept of energy, which one splits up into kinetic energy and potential energy; the equations of motion are derived from the *principle of least action* that had been postulated in a heuristic way by Maupertuis, and was formalized

in a rigorous way by Euler, Lagrange and Hamilton. The interest of this formulation of mechanics is due to its systematic nature: it provides a genuine methodology, comprising strict rules, which it is enough to observe rigorously to derive the equations of motion for any material system. As this methodology remains, in spite of certain adaptations and generalizations, at the heart of contemporary physics, it is worth taking some time to discuss its main concepts and moments.

A degree of freedom is a parameter, depending on time, that enters the definition of the position of a material object in space. A material point, for example, depends on three degrees of freedom, its three co-ordinates in a certain reference frame, and thus a system of N independent material points depends on $3N$ degrees of freedom. A fluid (liquid or gas) is a system depending on an infinite number of degrees of freedom, co-ordinates of the infinitesimal cells of which it is made up and that are comparable to material points. The state of a fluid can then be defined using one or several functions of these co-ordinates, which is called a *field*. As systems depending on an infinite number of degrees of freedom, fields can thus in principle be integrated into the program of mechanics. Let us note however that, at this stage, the concept of field is not a primitive concept: it is a secondary concept making it possible to account for the state of a given complex material system.

The state of a system depending on N degrees of freedom is represented by a single point, the coordinates of which, in an abstract space with N dimensions called the *configuration space*, are the N degrees of freedom. For such a system, the program of rational mechanics consists in determining, using the equations of motion and the initial conditions, the trajectory of the point representing the system in the configuration space.

The Lagrangian formulation of mechanics consists in making the equations of motion derive from a *variational* principle, known as the *principle of least action*. In mathematical terms, this principle stipulates that the trajectory followed in the configuration space by the point representative of a system is the one that minimizes a certain integral, called the *integral of action*, the integral over time of a function called the Lagrangian. This Lagrangian, which has dimensions of energy, is, for the simplest mechanical systems, equal to the difference between the kinetic energy and the potential energy.

The Lagrangian formulation of mechanics relies on the powerful variational method that consists in elucidating the dynamics of a physical process, in considering the whole set of ways the process can *virtually* follow and in establishing a criterion making it possible to determine the one *actually* followed.

Another advantage of the Lagrangian formulation is that it highlights particularly well the coordination between *relativity*, properties of *symmetry*⁶ and *conservation laws*. The Galilean principle of relativity is the true foundational principle of all mechanics, because it plays an essential role in allowing an objective approach of physical reality: are objective those aspects of reality that are maintained when one changes the reference frame, i.e. when one changes the point of view from which this reality is observed. Still it is necessary to define what "is maintained" when the change of reference frame takes place. One then has recourse to two narrowly connected concepts: on the one hand *invariance* (or *symmetry*), i.e. the fact that the

equations of motion do not change when one carries out certain transformations and on the other hand the *conservation* in the course of time of certain quantities. The Lagrangian formulation of mechanics makes it possible to establish a fundamental theorem, due to Emmy Noether, that mathematically gives an account of this coordination: with any property of relativity is associated a certain symmetry of the Lagrangian, i.e. a certain invariance of the Lagrangian with respect to certain transformations, and the law of conservation in the course of time of certain quantities. In mechanics, the theorem of Noether applies

- to the relativity of time, coordinated with the invariance with respect to time translations and the conservation of energy,
- to the relativity of space, coordinated with the invariance with respect to space translations and the conservation of momentum,
- And to the isotropy of space, coordinated with the invariance with respect to rotations and the conservation of angular momentum.

3 The Crisis of Mechanics

Thanks to the improvement of its formalism, analytical mechanics reinforced the hope that one can base on it a scientific conception able to account for the whole realm of observable physical phenomena. But in order for this prospect to take shape it was necessary to widen its field of application to phenomena that hitherto seemed to be foreign to it. The extensions of mechanics fall into two main categories with regard to its two basic concepts, the material point and the force. In connection with the concept of material point are the phenomena that could be integrated into mechanics thanks to the atomistic assumption, like heat phenomena, thermodynamics, and even chemistry. In connection with the concept of force are the electric and magnetic phenomena that the electromagnetic theory of light developed by Maxwell made it possible to associate with optical phenomena. Essentially, these extensions of mechanics were achieved by 20th century physics, but only at the price of completely restructuring its foundations.

At the beginning the crisis was signaled by a few very specific and academic problems, namely some phenomena that one was unable to quantitatively explain by means of the available mechanistic or mechanistically inspired models. To these puzzles belong the photoelectric effect that did not fit in the framework of Maxwell's theory of electromagnetism; the advance of the perihelion of Mercury, an effect disagreeing with the predictions of Newton's theory of gravitation; the specific heat of poly-atomic substances that challenged Maxwell's kinetic theory of matter, which aimed at unifying mechanics with the atomistic conception; the spectrum of black body radiation, which could not be described with the tools of thermodynamics and electromagnetism. The crisis was also fuelled by some unexpected experimental discoveries like those of X-rays by Roentgen in 1895, of the electron by Thomson in 1897, and of radioactivity by Becquerel in 1896 and Pierre

and Marie Curie in 1898. The discovery of radioactivity was the most intriguing one, since, although it suggested that atoms actually exist, it also suggested that they are not eternal and that they can undergo a change of species through a transmutation process.

In addition to the above-mentioned puzzles and discoveries, the program of rational mechanics was confronted with some conceptual questions that led it to a state of crisis. This crisis concerned the three domains of statistical, relativistic and quantum physics that we are going to review in the following sections.

4 Statistical Physics and the Problem of the Reality of Atoms

One can attribute to Carnot the foundation of theoretical thermodynamics: in an almost unnoticed work of 1824, *Reflexions on the Motive Power of Fire*,⁷ he makes the assumption that heat is a fluid, and starting from an analogy between the power of heat and that of a waterfall, he establishes what one can regard as the origin of the second principle of thermodynamics. To give rise to the power of heat, one needs a difference in temperature between a hot body and a cold body, and the output of any heat engine is necessarily lower than 1 (the maximum output is equal to the ratio of the difference in temperature to the highest temperature). But, in 1831, he questions the assumption of the heat fluid and a little further he states what is nothing but the first principle of thermodynamics (stated after the second one!), the principle of conservation of energy.

The complete formalization of thermodynamics is the work of Clausius who states in a clear way the two principles of thermodynamics: the first expresses the conservation of energy and the second one expresses, in terms of the increase of *entropy*, the impossibility of perpetual motion of the second kind (which would consist in producing work starting from only one heat source). The tendency of heat to pass in an irreversible way from hot bodies to cold bodies is explained by this second principle. After the work of Clausius, thermodynamics seemed a well-established theory, but its relations with mechanics were not clear. If energy seemed to lend itself to a mechanistic interpretation, other concepts of thermodynamics like pressure, temperature, or entropy did not seem to be easy to integrate into the framework of mechanics. It is thanks to the atomistic conception of matter and with the recourse to *statistical methods* that the synthesis of thermodynamics and mechanics took place through the *kinetic theory of matter* and *statistical thermodynamics* developed by Maxwell and Boltzmann.

The kinetic theory of matter made it possible, thanks to statistical methods, to determine some characteristics of the hypothetical constituents of matter called atoms or molecules, and to begin connecting the physical quantities of thermodynamics to the concepts of mechanics. A link is thus established between the microscopic laws of the elastic collisions of molecules and the first principle of thermodynamics, established at the macroscopic level, that of the conservation of energy. Temperature is interpreted in terms of molecular agitation: it is proportional

to the average kinetic energy of the molecules, namely half the product of their mass by the average value of the square of their velocity. The proportionality factor is Boltzmann's constant k . It is Boltzmann who completes the synthesis of thermodynamics and mechanics by establishing a mechanistic interpretation of entropy at the basis of the second principle: Boltzmann's constant acts as a proportionality factor between the entropy S and the logarithm of the number W of microscopic configurations, called *complexions*, giving rise to a given macroscopic state, $S=k\text{Ln}W$. Entropy thus gives a measure of the disorder that tends to increase with time for an isolated system, and the second principle of thermodynamics accounts for the fact that, insofar as randomness is at work, it is likely that a closed system presenting a certain order will go towards disorder, which offers so many more possibilities.⁸

In a conference intended for a wide audience, under the title "Molecules", Maxwell presented the recourse to the statistical methods as a makeshift to which we are constrained due to the imperfection of our means of knowledge and observation: "Thus molecular science teaches us that our experiments can never give us anything more than statistical information, and that no law deduced from them can pretend to absolute precision."⁹ To reach a world "where everything is certain and immutable", Maxwell said it is needed to pass "from the contemplation of our experiments to that of the molecules themselves, to leave the world of chance and change." This marks a severe conceptual difficulty: if molecules do exist, they are so small that they will never be observable and our knowledge about them will always be based on statistical assumptions, i.e. incomplete. This difficulty led some philosophers or physicists like Mach and Ostwald to adopt a positivistic stance and to reject the atomistic conception. The way out of this difficulty required, on one hand, providing statistical methods with a more solid theoretical ground and, on the other hand, discovering ways of making atoms or molecules experimentally observable.

At the very beginning of the 20th century, in 1902 precisely, it appeared, through the work of Gibbs and once again of the very young Einstein,¹⁰ that statistical methodology is not necessarily a makeshift but that its range is perhaps fundamental and universal. In the foreword of his *Elementary Principles of Statistical Mechanics*, Gibbs explains the major shift of point of view he proposes for the recourse to statistical methods:

We may imagine a great number of systems of the same nature, but differing in the configuration and velocities which they have at a given instant, and differing not merely infinitesimally, but it may be so as to embrace every conceivable combination of configuration and velocities; And here we may set the problem, not to follow a particular system through its succession of configurations, but to determine how the number of systems will be distributed among the various conceivable configurations and velocities at any required time, when the distribution has been given for some one time.¹¹

The advantage of so proceeding is that, as Gibbs says a little further:

The laws of statistical mechanics apply to conservative systems of any number of degrees of freedom, and are exact. This does not make them more difficult to establish than the approximate laws for systems of a great many degrees of freedom, or for limited classes of such systems. The reverse is rather the case, for our attention is not diverted from what is essential by the peculiarities of the system considered, and we are not obliged to satisfy

ourselves that the effect of the quantities and circumstances neglected will be negligible in the result.

The articles published by Einstein in 1902, without being acquainted by Gibbs' book proceed from his constant endeavor to work out the fundamental principles at work in a physical theory, specifically in statistical physics, the main object of his concerns at that time, while keeping as close a contact as possible with experiment. For Einstein, as for Gibbs, the concepts of statistical physics apply to ensembles of systems. Einstein considers ensembles that one calls today, following Gibbs, *canonical*, i.e. ensembles with a fixed temperature. Einstein also endeavors to transcend mechanics and to discover the most general statistical laws that do not depend on mechanistic modeling.

From the same thought process proceeds Einstein's endeavor to show that *fluctuations*, i.e. departures from the laws of thermodynamic equilibrium, able to affect "small systems" visible with the microscope, are accessible to experimental observation, which neither Boltzmann nor Gibbs believed. He presumes that the order of magnitude of these disturbances is related to Boltzmann's constant k and, since 1900, contemplates means of determining the characteristics of the atoms (their numbers, their sizes) using the observation of these fluctuations. In 1905 he succeeds in elaborating a theory of Brownian motion that could, in principle, be tested experimentally. When the existence of atoms was clearly established, after the experiments were carried out in accordance with this theory by Jean Perrin in 1908, this achievement was considered as a genuine triumph of rational mechanics, providing the scientific basis of the atomistic conception.

5 Restricted Relativity, Relativistic Particles and Fields

Another conceptual difficulty of rational mechanics is related to the controversy concerning the nature of light: is light made of waves or of corpuscles? This controversy was one of the subjects of concern to theorists at the end of the 19th century. True, Newton had proposed a corpuscular model for light, but the discovery of the phenomena of interferences and diffraction had tipped the scales on the side of an undulatory interpretation of light. The theory of the electromagnetic field developed by Faraday, Maxwell and Heaviside, strongly reinforced this interpretation when Hertz highlighted the fact that the waves of the electromagnetic field propagate precisely at the same speed as light: the propagation of light was then comparable with the propagation of waves of the electromagnetic field. But this conception raised difficult questions of a theoretical nature: one hitherto had never met waves which were not carried by a certain medium, or a certain fluid (it was known that there are no sound waves in the vacuum); what then was the medium "carrying the light waves"? One had thus postulated the existence of a mysterious fluid, called *ether*, which was supposed to carry the light waves. But then, such a medium was to be describable by means of rational mechanics, it was to induce observable effects, like

“an ether wind” due to the Earth moving in it. However all the theoretical and experimental efforts to establish the existence of this mysterious fluid appeared vain.

One can say that in 1905 the physics of electromagnetic interactions was in full crisis. The failure of the experiments of Michelson and Morley, aimed at testing the existence of an ether wind, was the subject of various interpretations. Independently of the model of ether, it came to be recognized that Maxwell’s equations are not invariant under the transformations known as Galilean, supposed to translate mathematically the principle of relativity, fundamental in mechanics: the laws of physics are expressed in the same way in two reference frames of inertia (i.e. in the absence of any external force) in relative rectilinear and uniform motion. It is Lorentz who discovered the transformations, called by Poincaré *Lorentz transformations*, and shown by him to form, together with spatial rotations, a *group*, which leave invariant Maxwell’s equations. However the significance of this invariance was not understood and its implications such as the contraction of length and the dilatation of time appeared very mysterious.

In 1905 Poincaré and Einstein produced almost simultaneously and independently their works on relativity. The work of Poincaré, founded on the Lorentz invariance of Maxwell’s equations, modeled the electron like an extended object, undergoing the “pressure of ether” in the form of a contraction in the direction of its motion. Einstein’s theory of relativity, which eliminates the very idea of ether, is very different: it affects the most fundamental part of mechanics, namely kinematics, the very doctrine of space and time. Einstein first shows that, because of the finite time that light (or any other signal possibly carrying information) puts to be propagated, it is impossible to decide in an absolute way of the simultaneity of two instantaneous events spatially separated. He thus reinterprets the speed of light in the vacuum c as a universal constant translating the absence of instantaneous interaction, and he redesigns mechanics by adding to the principle of relativity the principle of the *invariance of the speed of light*. This refoundation implies that one abandon the absolute character of time (two clocks in relative motion do not mark the same time) and the absolute character of spatial metric (two identical rulers in relative motion do not measure the same length). According to the expression suggested some time afterwards by Minkowski, time in this new kinematics must be regarded as the fourth dimension of *space-time*, a continuum whose other three dimensions are those of space. In this new kinematics, the Lorentz transformations express the way in which space-time co-ordinates change in a uniform rectilinear motion with a speed necessarily lower than or equal to the speed of light. A little time after this historic article, again in 1905, Einstein established the principle of the inertia of energy, which is translated in his most famous formula $E=mc^2$. A material point of mass m , moving in a rectilinear uniform motion has an energy E and a momentum \mathbf{p} that form a 4-vector of space-time, called the four-momentum (the analogue of a 3-vector in the three-dimensional space of classical mechanics). Einstein’s famous formula is a particular case of a relation between the mass, the energy and the momentum, known as the *mass shell or dispersion relation* which expresses the fact that the norm of the 4-momentum (the analogue of the length of a 3-vector), equal to mc^2 , is invariant under the Lorentz transformations (in the same

way as the length of a 3-vector is invariant under space rotations). In a space-time reference frame where the material point is at rest, namely where its momentum vanishes, the norm of the 4-momentum reduces to the rest energy or proper energy, which thus equals mc^2 . This relation between mass and energy is a true novelty of relativity: in classical mechanics a particle at rest has no energy since the only energy that a material point may have is its kinetic energy which vanishes when the velocity vanishes, whereas, in relativity, even at rest, a particle has a proper energy, that, in units in which the speed of light is a large number, is enormous. It is interesting to note that the mass shell relation allows the value zero for the mass, which is also a novelty with respect to classical mechanics, for what could a material point of zero mass mean? In relativity a mass-less particle is never at rest, it moves, just as light, at the speed of light in any reference frame; it has an energy and a momentum equal to the energy divided by c . In a sense one could say that, whereas in classical mechanics mass precedes energy (there is no energy without mass), in relativity energy precedes mass (there is no mass without energy).

With this relativistic kinematics, implying the Lorentz invariance for all phenomena, it becomes possible to integrate the electromagnetic theory within the renewed framework of mechanics. In this framework the new fundamental concept is the concept of *field*, of which the electromagnetic field is an archetype. A field is a physical object, with an infinite number of degrees of freedom, extended to the whole of space-time: it corresponds to the definition, at each point of space and at any instant of time, of a function or a set of a few functions. So conceived, the electromagnetic field does not need unspecified ether or any carrying medium; it is itself the seat of the oscillatory phenomena associated with the propagation of light. The electromagnetic field carries energy and a momentum equal to the energy divided by c , so one can say that it is a *mass-less field*. A particle like the electron has a specific property, called its electric charge, which makes it able to produce an electromagnetic field and to react to the action of such a field. The electromagnetic interaction is not propagated instantaneously at a distance: a moving charged particle produces a variable electromagnetic field, the variations of which can subsequently put in motion another particle spatially separated from it.

6 General Relativity, Gravitation and Cosmology

Einstein then seized this concept of field and tried to make it into the most fundamental concept of the whole of physics. His research then went on to generalize the theory of relativity. Not seeing any reason that the principle of relativity should be restricted to the changes of inertial reference frames, he sought to extend this principle to the most general changes of reference frames. He succeeded in reaching that aim thanks to a detour through the theory of gravitation: by noting that the acceleration produced by gravitation on a material body does not depend on the mass of this body, he showed that a change of reference frame comprising acceleration is equivalent to a gravitational field of opposite acceleration. More generally,

he established that any change of reference frame can, locally, be replaced by a certain gravitational field, and that, reciprocally, any gravitational field can, locally, be replaced by a certain change of reference frame. In this sentence, the adverb *locally* means that the equivalence between the gravitational field and the change of frame is only possible in an infinitesimal region of space-time. Applied to the propagation of light, this reasoning implies that light undergoes the action of gravitation, which, we recall, is acceleration. To safeguard the invariance of the speed of light, Einstein was led to postulate that the effect of gravitation is a modification of the metric of space-time: gravitation influences the length of the measuring-rods and the running of the clocks, in such a way that the speed of light remains constant! Thus the generalization of the theory of relativity leads to a new theory of universal gravitation, geometrical in nature: *matter and the gravitational field that it induces are replaced by a space-time the metric of which is a universal field*. In 1915, Einstein put into equation this masterpiece, in terms of a theory of universal gravitation, which encompasses that of Newton, reduces to it at the approximation of weak fields, makes it possible to solve the puzzle of the motion of Mercury's perihelion, and finally, predicts new effects, such as the deflection of light by heavy stars, which was observed during the solar eclipse of 1919.

Immediately after having elaborated the theory of universal gravitation based on general relativity, Einstein tried to apply it to cosmology. He first noticed that Newton's theory of universal gravitation is not in harmony with the observation that the density of matter in the universe is in average approximately uniform, whereas it predicts rather a maximum density of stars at a sort of a center and decreasing to zero far away from this center, "a stellar universe [that] ought to be a finite island in the infinite ocean of space."¹² He then showed that thanks to the non-Euclidean character of the geometry implied by general relativity, one can conceive a universe that is finite and yet without boundary.

7 Relativity and the Problem of Space

The title of this section is taken from the fifth appendix added by Einstein in 1952 to the fifteenth edition of his book *Relativity*, in which he had explained, as early as 1917, restricted and general relativity for a wide audience. In this appendix he expresses the wish "to show that space-time is not necessarily something to which one can ascribe a separate existence, independently of the actual objects of physical reality,¹³" and that finally "the concept of 'empty space' loses its meaning". In this very dense text, Einstein exposes his epistemological views about space and time. To conceive physical reality one needs the concept of *event* and the concept of *material object*. He first notes that "it is just the sum total of all events that we mean when we speak of the 'real external world'" and then that "it appears to [him], therefore that the formation of the concept of the material object must precede our concepts of time and space." He goes on to discuss the evolution of the conception

of matter, space and time from classical Newtonian mechanics to restricted and general relativity. In Newtonian mechanics physical reality:

thought of as being independent of the subject experiencing it, was conceived as consisting, at least in principle, of space and time on one hand, and of permanently existing material points, moving with respect to space and time, on the other; The idea of the independent existence of space and time can be expressed drastically in this way: If matter were to disappear, space and time would remain behind (as a kind of stage for physical happening).¹⁴

The passage from classical mechanics to restricted relativity is characterized by the promotion of the concept of field that “becomes an irreducible element of physical description, irreducible in the same sense as the concept of matter in the theory of Newton.” However this evolution in the physical description does not affect the idea of the existence of space (more precisely this space together with the associated time) as an independent component of the representation. Also, even when they have been made compatible with restricted relativity, the electromagnetic theory and the rest of mechanics still need the concept of material points, possibly carrying electric charges. In the general theory of relativity, the concept of field acquires a more important status, because, on the basis of this theory:

Space, as opposed to ‘what fills space’, which is dependent of the co-ordinates, has no separate existence. [...] If we imagine the gravitational field [...] to be removed, there remains absolutely nothing. [...] there exists no space ‘empty of field’.¹⁵

8 Quantum Physics: From the Discovery of the Quantum of Action to Quantum Mechanics

The introduction of the elementary quantum of action by Planck in 1900 in his formula accounting for the spectrum of black body radiation initiated a long period of research and strong controversies that led to the current universal agreement about the fundamental status of quantum physics. True, the implications of the quantum of action were very intriguing: as soon as agreement was reached concerning the undulatory interpretation of light, one discovered, through Planck’s formula and its interpretation by Einstein in terms of energy quanta, that it has also a possible corpuscular interpretation; as soon as it was possible to clearly reject the positivistic objections against the atomistic conception, one discovered that, because of their quantum properties, atoms cannot be thought of as material points. More fundamentally, as an element of discontinuity in action, Planck’s constant and the physics in which it enters put the crisis of mechanics at a genuine climax, because it questions the two pillars of the whole scientific enterprise, namely, *causality* and *objectivity*. Causality is questioned because, in classical mechanics, as we said above, the causal laws of motion are derived from a principle of least action, which imperatively requires the continuity of action, and one does not know how to apply it if there is an elementary quantum of action. Objectivity is

also questioned since, at the quantum level, the object to be observed is modified, transformed by the observation. If one wants to observe a microscopic structure with a high spatial and temporal degree of accuracy (i.e. with a small spatial and temporal margin of error), it is necessary to transfer to it, for a certain length of time a certain quantity of energy. The product of this duration by this energy has to be at least equal to Planck's constant. But since the duration of the measurement must not exceed the tolerated temporal margin of error, the energy necessary for obtaining a result of measurement will be at least inversely proportional to this temporal margin of error. True, this circumstance does not bear any consequence as long as one remains in the field of classical physics, i.e. when the actions brought into play are very large with respect to the elementary quantum of action, but as soon as one wants to explore with sufficient precision the atomic or subatomic world, it obliges us to give up the implicit prejudice according to which it is always possible, at least in principle, to disregard the condition of observation: in its preparation, as well as in its results, any experiment in the microscopic world depends in such an essential way on these conditions that they must be taken into account down to the very formalism itself. Such a constraint seems to question the possibility of an objective description of the microscopic world.

The resolution of such a crisis took about 30 years of trials and errors, controversies, new experimental discoveries and conceptual innovations to lead to what came to be called *quantum mechanics*, comprising a rigorous mathematical formalism and a physical interpretation. Although the discovery of the quantum of action took place in the field of electromagnetic radiation, a field not directly related to mechanics, and although the contributions of Einstein, till the mid 20's mainly concerned the quantum theory of radiation, the founders of quantum physics concentrated on "quantizing" non-relativistic mechanics of point particles, postponing for a further stage the quantization of (relativistic) field theory.

The formalization of quantum mechanics was carried out at a frantic rhythm in 1925 and 1926. It is initially Heisenberg who, in 1925 and in collaboration with Born and Jordan, developed a completely new approach that was called the *mechanics of matrices*, which associates with the observable physical quantities matrices obeying relations of commutation. On his side, P. Dirac arrived by a different way of thinking to a formalization of what he called *quantum mechanics*, (the title of the thesis he defended in 1926). It is likewise in 1926 that Schrödinger developed, with the aim of making comprehensible the wave-corpuscule duality of de Broglie, a third approach, called *wave mechanics*, based on the *wave function* that obeys the now celebrated *Schrödinger's equation*. Some time later, again in 1926, Schrödinger showed the equivalence of his approach with that of Heisenberg, as well as that of Dirac. A coherent formalism, primarily founded on Schrödinger's equation, thus began to emerge, which made it possible to account in a precise way for the experimental observations like, for example, the Stark and Zeeman effects.

To these advances in the formalization, it is worth adding two major contributions pertaining to interpretation: the probabilistic interpretation of the wave function suggested by Max Born in June 1926, and the principle of indeterminacy stated by Heisenberg in 1927. Thus, at the end of the 20's, a consensus was reached on a

formalism and an interpretation, known as the *Copenhagen interpretation*, which made it possible to elucidate the problems left open by classical physics and to undertake the systematic exploration of the quantum universe.

Although it is called mechanics, the physics that quantum mechanics is supposed to describe has several features that seem completely foreign to rational mechanics. A first such feature is the *particle-wave duality*. Whereas the observation of the Compton Effect confirmed the existence of a corpuscular structure in the electromagnetic field that hitherto was conceived only in an undulatory way, Louis de Broglie, proposed, in his PhD thesis in 1924, that corpuscles of matter, like electrons, can show undulatory aspects. These ideas were confirmed by the observation of the phenomenon of interferences and diffraction induced by electrons. “The work of de Broglie made me a great impression. It lifted a corner of the great veil¹⁶” said Einstein, impressed by this vision. Gradually, it indeed appeared that in the quantum world (i.e. when the actions involved are of the order of magnitude of the elementary quantum of action) both in the realm of the structure of matter, and in the one of the interactions, phenomena are suitable for two descriptions, which would be completely contradictory in the framework of classical physics, one in terms of waves and another in terms of particles. The frequency and the wave vector that characterize the propagation of the wave are proportional to the energy and the momentum that characterize the motion of the particle with a proportionality factor equal to Planck’s constant.

Another very intriguing feature of quantum mechanics is the *superposition principle*. Whereas, in classical mechanics, the states of a system are represented by points of the space of configuration, they are represented, in quantum mechanics, by vectors of a *Hilbert space*, a linear vector space of complex functions on which are defined a norm and a scalar product. One also uses the term of *wave function* to indicate a vector of the Hilbert space representing a quantum state. The linearity of the Hilbert space corresponds to the superposition principle according to which quantum states can combine, superimpose, i.e. can be added like complex numbers, as do, in classical physics, waves or fields. This property of *coherence* is one of the essential characteristics of the entire quantum universe. But it is also this property which is at the origin of the most disconcerting and paradoxical aspects of this new physics: one could thus imagine thought experiments in which a physical system could be in a state of superposition of two contradictory states (as the poor cat which Schrödinger had imagined, at the same time dead and alive).

9 Einstein’s Criticism of Quantum Mechanics

Another essential characteristic of quantum mechanics that is revealed by radioactivity is that its predictability is *probabilistic*. One is obliged to resort to probabilities, on the one hand because there are processes, bringing into play an action of the order of the elementary quantum of action, like a radioactive decay or a nuclear or particle reaction, which it is impossible to describe in a deterministic

way using differential equations, and on the other hand because it is necessary to include in the formalism the conditions of observation and that these conditions cannot in general be better determined than in a statistical way. This feature was Einstein's main concern in his criticism of quantum physics. His attitude towards quantum physics varied with time. Till the mid 20's, not only did Einstein not criticized quantum physics but, as one of its founders he very warmly praised the advances it made possible. A careful reading of his articles shows that what he tries to establish is a quantum theory of fields rather than a quantum mechanics: this appears in his 1905 article on the photoelectric effect and in his famous article in 1917, "Quantum Theory of Radiation", in which he provides a demonstration of the Planck formula for the black body radiation; even in his articles in 1924 and 1925 on the quantum theory of the mono-atomic ideal gas, in which he integrates the Bose statistics (now known as the Bose–Einstein statistics) in the framework of quantum physics, he notes that "it is possible to associate a field of scalar waves with a gas." In any case, the feature that he never accepted is the recourse to probabilities at the fundamental level, because this recourse would imply that the *theory is incomplete*. In his "Reply to Criticisms", quoted above, he considers a radioactive decay described in quantum mechanics by means of a "Psi-function" (i.e. a wave function):

This Psi-function yields the probability that the particle, at some chosen instant, is actually in a chosen part of space (i.e., is actually found there by a measurement of position). On the other hand, the Psi-function does not imply any assertion concerning the time instant of the disintegration of the radioactive atom. Now we raise the question: Can this theoretical description be taken as the complete description of the disintegration of a single individual atom? The immediately plausible answer is: No. For one is, first of all, inclined to assume that the individual atom decays at a definite time; however, such a definite time-value is not implied in the description by the Psi-function. If, therefore, the individual atom has a definite disintegration time, then as regards the individual atom its description by means of the Psi-function must be interpreted as an incomplete description. In this case the Psi-function is to be taken as the description, not of a singular system, but of an ideal ensemble of systems. In this case one is driven to the conviction that a complete description of a single system should, after all, be possible, but for such complete description there is no room in the conceptual world of statistical quantum theory.¹⁷

In the celebrated "EPR" *Physical Review* paper, written in 1935 in collaboration with Boris Podolsky and Nathan Rosen, "Can Quantum-Mechanical Description of Physical Reality be Considered Complete¹⁸?", Einstein proposes a thought experiment that could lead to a paradox possibly ruining the whole consistency of quantum physics. In this paper, the paradox was formulated by means of a pure thought experiment concerning the determination of the positions and momenta of a pair of particles produced in a well-defined quantum state. Although, for each particle of the pair, the position and the momentum obey the law of non-commutation and can thus be determined only with uncertainties constrained by the inequalities of Heisenberg, the difference of the positions commutes with the sum of the momenta. It would thus seem that one could measure with an arbitrarily high precision this difference and this sum and that consequently one could predict with precision either the value of the position or that of the momentum of the first particle of the pair, if,

respectively, the value of position or that of momentum of the second particle of the pair is measured. Since, at the time of measurement, the direct interaction between the particles of the pair has ceased, the position and the momentum of the first particle can be regarded as physical attributes of an isolated object, which would mean that one could “beat the inequalities of Heisenberg”, and thus that quantum mechanics does not provide a complete description of reality.

In a letter to Schrödinger of June 19th 1935, Einstein reconsiders the EPR thought experiment of which he presents the implications in the form of a true antinomy: either quantum theory is incomplete or it violates what he calls a *separation principle* according to which if one considers a system whose real state is composed of the real states of two subsystems A and B, then the real state of subsystem B cannot depend in any way on the experiment one performs on subsystem A.

The complete elucidation of the EPR paradox took several years. It required several advances on the experimental and theoretical grounds. A first advance was made by David Bohm, who imagined possible experiments, more realistic than that evoked in the EPR article, in which the position and the momentum are replaced as non-commutative observables by components of spins on different axes, which, in quantum mechanics, are represented by operators which do not commute. On a theoretical grounds, it is John Bell who, in 1964, established some inequalities that should satisfy the results of the experiments imagined by Bohm, on the first assumption that quantum mechanics would be incomplete and would thus have to be supplemented with some *hidden variables* and on the second assumption of *locality* i.e. the assumption of absence, in accordance with Einstein’s principle of separation, of an instantaneous connection between spatially separated systems. These inequalities thus made it possible to put Einstein’s argument to a precise quantitative test: either they would be satisfied, and then Einstein would be right, or they would be violated, and then at least one of the two assumptions made by Bell (hidden variable or locality) would be at fault. In the 70’s, some experiments aiming to test the Bell’s inequalities were carried out in atomic physics and nuclear physics, but it is in 1982 that the decisive experimental advance occurred: Alain Aspect and his collaborators succeeded in carrying out a genuine EPR experiment (in the version of a Bohm experiment); they found, and this was confirmed by many other experiments carried out since, a clear violation of Bell’s inequalities, thus confirming the predictions of quantum theory.

10 Mature Quantum Physics, the Quantum Theory of Fields

With the failure of the lawsuit in incompleteness brought by Einstein against quantum physics, the verdict of the experiment is without appeal: quantum physics is discharged. Therefore, in at least one of his criticisms, Einstein was wrong. With the encompassing view that more than 70 years of implementation of quantum physics give, it is advisable to reassess the objections he made to this physics, to locate in what respect he was right and in what respect he was wrong, and also to evaluate, in a critical way, the Copenhagen interpretation to correct its possible defects.

We believe that it is the passage from quantum mechanics to the quantum theory of fields that enables us to answer the epistemological objections raised by Einstein with respect to locality, reality and completeness, and thus to solve the crisis of physics initiated by the discovery of the elementary quantum of action.

Not only was Einstein entirely right to require what he called the principle of separation, but one can blame the Copenhagen interpretation for not having sufficiently stated it. Expressed bluntly by Steven Weinberg, who calls it the *cluster decomposition principle*, in his textbook on the quantum theory of fields, it affirms that

Experiments that are sufficiently separated in space have unrelated results. The probabilities for various collisions measured at Fermilab should not depend on what sort of experiments are being done at CERN. If this principle were not valid then we could never make any predictions about any experiment without knowing everything about the universe.¹⁹

This principle, also called the *principle of locality*, indeed seems to be one of those with which it is really impossible to compromise.

Several of the Einstein's queries about quantum physics are related to the question of reality: the belief in the existence of a material reality, independent of any observation, and describable in space and time; the difficulty in defining what is "reality" since it is known to us only by the description that physics gives it; the dualism of the field and the material point, two descriptions that are possible but contradictory. This dualism, which Einstein always rejected and he was unable to get rid of, is indeed overcome by the quantum theory of fields, as Weinberg says in an article, under the title "What is Quantum Field Theory and What did We Believe It Is?" in which he highlights some topics of his textbook:

In its mature form, the idea of quantum field theory is that quantum fields are the basic ingredients of the universe, and particles are just bundles of energy and momentum of the fields. In a relativistic theory the wave function is a functional of these fields, not a function of particle coordinates. Quantum field theory hence led to a more unified view of nature than the old dualistic interpretation in terms of both fields and particles.²⁰

To address the question of completeness, we need to go back to the above-mentioned articulation of the two basic concepts necessary to conceive reality, the concept of *object* and the concept of *event*. The concept of object belongs to the realm of theory, whereas the concept of event belongs to the realm of experiment: the aim of theory is to constitute a scientific object, an element of reality independent of the way it is observed; events on the other hand are the modalities through which reality is empirically or experimentally known to us. Completeness is a *theoretical* requirement, not an experimental requirement that thus concerns the object not the event. On the one hand, Einstein was right when he blamed quantum mechanics to keep the particle as a representative of the primitive concept of object while giving the wave function a probabilistic (i.e. incomplete) interpretation; but, on the other hand he was wrong in his hope that quantum events be individually predictable in a deterministic way. The finiteness of the elementary quantum of action forbids any subdivision of individual quantum processes. These processes must be considered as irreducible events that are neither individually predictable nor reproducible. In the framework of relativity, general covariance requires events to be strictly localized in space-time. In the quantum

framework, even in absence of relativistic effects, it is the principle of locality that requires quantum events to be strictly localized in space and time. The only possible predictability concerning quantum processes is probabilistic by means of statistical averages over ensembles of strictly localized events occurring in some region of space-time.

A quantum field is a physical entity defined at each position in space and instant in time. Whereas a classical field entity is a real or a complex function of the space-time coordinates, a quantum field entity is an operator that produces or destroys a particle in a quantum event strictly localized at the space-time coordinates. According to the quantum theory of fields the particle-wave duality is interpreted in a non dualistic way: quantum fields are *objects* that behave, either as particles or as waves according to their being involved or not involved in actual *quantum events*. As Feynman says in the article in which he introduced the *path integral* reformulation of quantum physics,

The electron acts as a wave, so to speak, as long as no attempt is made to verify that it is a particle; yet one can determine, if one wishes, by what route it travels just as though it were a particle; but when one does that [the classical way to combine probabilities] applies and it does act like a particle.²¹

Quantization of field theory is often named “second quantization”. According to this terminology, the first quantization is the association with a system of particles of a wave function that is considered as a classical field, the quantization of which is the second quantization. Actually, it appears that the quantum theory of fields is rather a complete change of perspective. In quantum mechanics, the *states of the system* are represented by *vectors* of the Hilbert space, and the *observable physical quantities* are represented by *operators* acting on these vectors. In quantum field theory there is a complete change of point of view: operators are associated with the object, the quantum field, whereas vectors are associated with the states, not of the system, but rather of the experimental apparatus. A quantum field operator that produces or destroys a particle acts on the state of the particle detector. In quantum mechanics, the wave function of a particle is a complex function of the space and time coordinates, or of the energy and momentum, the squared modulus of which is the probability that *the* particle has these coordinates or these energy and momentum. On the other hand, according to the quantum field theoretical point of view, the wave function is a field amplitude, a complex function, the modulus squared of which is the probability of counting at the corresponding position or with the corresponding energy and momentum *a* particle produced by the quantum field. Actually, it turns out that in quantum physics, all experiments are more naturally interpreted according to this quantum field theoretical point of view than according to the quantum mechanical point of view, for all the detectors that enable us to experimentally observe the atomic or subatomic world are nothing but *event counters*, possibly including some filters that make it possible, say, to select particles with a given spin component, but never apparatuses that would enable us to determine the wave function of an isolated particle. Having this in mind, one understands why the passage from the quantum mechanical point of view to the quantum field theoretical point of view provides

a solution to the EPR paradox: as Einstein himself noticed, there is no paradox if experiments are interpreted in terms of statistics of ensembles. The only mistake Einstein made was to consider such ensembles as ensembles of systems and not as ensembles of events.

11 Quantum Field Theory and the Physics of Fundamental Interactions

Historically, quantum field theory was applied for the first time in Quantum Electrodynamics (QED), that is the quantum field theory of the electromagnetic interactions of electrons and positrons. It is in order to work out a tractable scheme suited for this purpose that Feynman was led to elaborate his above-mentioned *Path Integral Quantization* as an alternative to the standard methods of quantization that were available at that time. Starting from the simplest example, i.e. the quantum mechanics of a one-particle system, he rewrites the Schrödinger's equation as a functional integral equation, the solution of which, the wave function of the particle at a given space-time position, is a functional integral (that is an infinite dimensional integral) over all the "paths" or trajectories that could possibly bring the particle from an arbitrary position in the infinitely remote past to its actual position. Such a reformulation looks very complicated in the very simple case considered, but it can be applied to very general situations, including the treatment of fundamental interactions with quantum field theory. The integrand of the path integral, namely the weight given to the contribution of each path (in the case of a field theory, one should rather speak of each "field history") involves the Lagrangian of the theory in which is encoded all the information concerning the considered interaction (the fields involved, the masses, spins and other quantum numbers of their quanta, the symmetries of the interaction, the *coupling constants* that characterize the intensity of the interaction at the elementary level, etc.) The Lagrangian is the sum of the kinetic energy terms corresponding to the free propagation of the fields involved and of the interaction terms corresponding to the interactions or couplings of the fields. The locality principle constrains all the terms in the Lagrangian to be of the form of products of fields, or field derivatives *evaluated at the same space-time point*.

The quantization of field theory confronted two major difficulties, negative energies and infinities, the overcoming of which is one of the keys of the success of the standard model.

The first difficulty arose as soon as one tried to work out a relativistic generalization of the Schrödinger's equation. Even for free particles, in which case standard and path integral quantization can be worked out explicitly and lead to the same results, such a generalization leads to negative energy solutions that would imply that no quantum state could be stable since the energy would not be bounded from below. The physical interpretation of these negative energy solutions is impossible in the framework of quantum mechanics where the number of particles is fixed and

conserved. It is precisely the passage from quantum mechanics to the quantum field theory that makes it possible to overcome this difficulty: in quantum field theory the number of particles is not conserved; particles can be produced or destroyed, and the problem of negative energies is solved by constraining negative energy, i.e. unphysical particles to *go backward in time* and replacing such a negative energy particle with a given charge by a positive energy, i.e. a physical *antiparticle* with the opposite charge *going forward in time*. With this scheme time is axiomatically given an arrow: *only physical particles or antiparticles go forward in time*. The experimental discovery of the *positron*, the antiparticle of the electron, and then of the antiparticles of all the known particles has clearly demonstrated the adequateness of this scheme.

When interactions are taken into account, the standard quantization methods lead to very cumbersome, almost intractable calculations, whereas path integral quantization leads to a very powerful scheme known as the *perturbative expansion* in terms of *Feynman's diagrams and amplitudes*. For any process relying on a given fundamental interaction, the amplitude the modulus squared of which is the probability of its occurrence, can be expanded in powers of the coupling constant, the coefficients of which are a sum of Feynman's amplitudes associated with Feynman's diagrams. These Feynman's diagrams make it possible to picture in a very suggestive way the basic idea of the path integral of decomposing an *actual* process in terms of a sum of terms associated with *virtual* processes. The higher the power of the coupling constant in the power expansion is, the more complex are the Feynman diagrams, so, if the coupling constant is a small number (as it is the case in QED) one can hope to get with the contributions of a few simple virtual processes a good approximation of the full amplitude.

The amplitude associated with a Feynman diagram is always written in terms of multiple integrals over a finite number of variables. At this point one has to confront the difficulty of infinities: in general the integrals necessary to compute Feynman's amplitude *diverge*, namely are equal to infinity. Actually, this difficulty, which seems to possibly ruin the entire quantum field theoretical program, is deeply rooted in the conflict, already raised by Einstein, between locality and completeness: locality requires considering point-like couplings of fields, which in turn requires taking into account virtual processes involving arbitrary large energies responsible for the divergent integrals; if, in order to get finite amplitudes one would simply ignore the virtual processes involving energies higher than some arbitrary *cut-off*, then the theory might be blamed for incompleteness. The idea of a way out of this difficulty is to split the values of the parameters of the theory into their *bare values*, i.e. the values they would have in absence of interaction, and their *physical values*, i.e. the values they acquire due to the interactions. In QED, where the parameters are the electron mass and the electron charge, it turns out that infinities arise when one tries to express the physical amplitudes in terms of the bare values of the parameters whereas no infinity occurs in the expression of the amplitudes in terms of the physical values of the parameters. Of a theory, like QED, in which such a "miracle" occurs for all amplitudes and at all orders of the perturbative expansion, one says that it is *renormalizable*. Since the physical values of the parameters can be experi-

mentally determined, it is possible to compare with experiment the predictions of a renormalizable theory. In the case of QED, for some physical quantities that are theoretically calculable and experimentally measurable, the agreement between theory and experiment is amazingly good.

12 Towards a Philosophical Category of *Reality Horizon*

From the rational explanation of this “miracle” can be drawn the main philosophical lesson of the present chapter. Actually, the physical values of the parameters implicitly depend on an energy associated with the *coarse graining* with which the interaction is experimentally observed. The realization of this coarse graining dependence is an asset of what is known as the *modern interpretation* of quantum physics that, in turn, is an asset of the path integral quantization method. In order to be able to attribute probabilities to actual events produced by interacting quantum fields one has to perform the path integral with a graining that is sufficiently coarse so that interferences that might prevent ascribing additive probabilities to independent events actually cancel.²² Now, because of that circumstance, a renormalisable theory like QED cannot be considered as a fundamental theory valid at all energies, but rather as an *effective* theory, suited to describe the interaction at a given resolution related to the coarse graining energy. But does that not imply such a theory to be incomplete since it would depend on parameters varying with energy? Actually, this is not the case because the way in which the parameters depend on the coarse graining energy is not arbitrary: it has to be such that the measured and calculated physical quantities *do not depend on it*. The equations that translate this physical independence on the coarse graining are called the *renormalization group equations*. According to the QED renormalization group equations, the fine structure “constant”, equal to the square of the electron charge divided by the product of Planck’s constant by the speed of light is not constant: it is *predicted* to vary from 1/137 at an energy of a MeV (a million electron-Volt) to 1/128 at an energy of a hundred GeV (a hundred billion electron-Volt), and this prediction has been confirmed by experiment. On the physical ground, the great achievement of the standard model is that one has embedded QED in a set of renormalizable theories (the electroweak theory and Quantum Chromodynamics, QCD) leading to predictions that have been experimentally confirmed with an excellent accuracy.

On an epistemological ground these achievements have put in the foreground a concept that currently plays a growing role in the context of quantum cosmology, the concept of *horizon*. In contemporary physics this concept is relevant in the interpretation of the fundamental limitations of human knowledge implied by some universal constants²³ like Planck’s constant or the velocity of light: these limitations are not to be considered as insuperable obstacles but rather as *informational horizons*, namely some boundaries beyond which lie some inaccessible information. The fact of assuming the existence of an informational horizon does not mean that one neglects or forgets the information lying beyond it. The method-

ology that allows keeping track of this missing information is based on functional integration: to evaluate the probabilities of the values of the dynamical variables bearing the accessible information (the followed variables) one *integrates out* the dynamical variables bearing the inaccessible information (the non-followed variables). Such a methodology is used in classical statistical physics where the microscopic configurations leading to the same macroscopic state (what one calls the *complexions*) are treated as non-followed variables that are integrated out through the statistical averages leading to the definition of the Boltzmann–Gibbs probability distribution of the followed variables. Basically, the path integral quantization relies on the same methodology: the summation, with a certain coarse graining, over all possible paths or field histories exactly corresponds to integrating out non-followed variables. Actually, it can be shown that the similarity between the Boltzmann–Gibbs probability distribution in statistical classical physics and the path integral in quantum physics is not a simple analogy, but rather a rigorous mathematical correspondence, with a strict “dictionary” translating Boltzmann’s constant into Planck’s constant, entropy (or information) into action, inverse temperature into imaginary time, critical phenomena occurring at a second order phase transition into the results of a renormalisable quantum field theory. The last item of this dictionary led in the 70’s to a remarkable interdisciplinary synthesis, since one was able to use, with great success, the same theoretical tools in two domains of physics which hitherto seemed completely disconnected, the physics of critical phenomena on one hand and Quantum Chromodynamics, the physics of strong interactions of quarks and gluons on the other. In this respect it is interesting to note that the same correspondence allowed designing some computer simulations of QCD, the so called “lattice QCD” providing some insight on the non-perturbative regime of this quantum field theory.

A last comment is in order about the correspondence between classical statistical physics and quantum physics. Since an imaginary time can be considered as a fourth Euclidean dimension of space, one can say that somehow quantization adds an extra space dimension to classical physics: quantum physics in a three-dimensional space is equivalent to classical statistical physics in a four-dimensional space. Such a feature is analogous to what occurs in the reconstruction of a three-dimensional scene by means of a two-dimensional *hologram*. Following this line of thought, some very important developments currently occur in cosmology. Gravitation is the only interaction capable of so much curving space-time that it leads to the formation of a spatial horizon, namely a “one-way membrane”, a two-dimensional informational horizon hiding information lying beyond it. Because of the expansion of universe, there exists in cosmology a horizon, called the *particle horizon* that is defined by the distance beyond which lie galaxies whose light had not the time to reach us. Beyond that horizon one suspects the existence of another horizon, called the *event horizon* that would be defined by the distance beyond which no information can *ever* reach us. This event horizon is usually assumed to rely on *quantum cosmology*, i.e. the domain of cosmology in which gravity has to be quantized. A theoretical laboratory to explore the physics of such event horizons is the physics of *black holes*. The event horizon of a black hole is the surface surrounding it beyond which any mat-

ter (and thus any information), trapped by the black hole escapes from perception. Although black hole physics is classical as far as gravitation is concerned, at the horizon, the classical gravitational field is so intense that it may induce in matter certain quantum effects such as the production of particle–antiparticle pairs, which have to be dealt with. Since, in quantum statistics, missing information is equivalent to entropy, it is natural, in this framework, to attribute entropy to such a horizon. Bekenstein and Hawking have shown that the entropy corresponding to the information trapped inside a black hole is proportional to the area of the event horizon rather than to the volume embedded inside it. It seems possible to generalize this result to space-time metrics involving a horizon which leads to conjecture that cosmology associated with such metrics is completely determined by the quantum properties of the horizon.²⁴ According to such a *holographic principle*,²⁵ the total information contained in a universe involving a horizon would not be proportional to the volume embedded by the horizon but only to the area of the horizon.

On a philosophical ground, I would like to conclude this chapter by emphasizing the relevance to philosophy of science of a concept that could act as a genuine philosophical category, the concept of *reality horizon*. The reality horizon is one of the key concepts of the philosophy of Ferdinand Gonseth (1890–1975), a Swiss mathematician-philosopher who was familiar with theoretical physics (he was a close friend of Michele Besso,²⁶ the closest friend of Einstein; he was asked by Georges Lemaître, one of the founders of modern cosmology, to write a foreword for his book *The Hypothesis of the Primitive Atom*²⁷) and who designed what I think is the philosophy that 20th century science deserves. In a development in his major book *Geometry and the Problem of Space*, devoted to the articulation of the three essential aspects of geometry, namely intuition, axioms and experiment, he notes that

The previous results have a value that goes beyond the framework of geometry. They concern the entirety of knowledge, we mean the state in which knowledge comes to us, at a given instant: Nothing authorizes us to think that our knowledge, even at its last frontiers, is more than a knowledge horizon; that the last ‘realities’ that we have conceived are more than a reality horizon.²⁸

It seems to me that all the developments of 20th century physics, from the resolution of the crisis of rational mechanics to the promising speculations about quantum cosmology through the successes of the standard model, confirm the validity of this ambitious and yet humble philosophy: we are such, and the world is such that it is never given to us in its full reality but as a reality horizon.

Endnotes

- 1 For a text-book intended for physicist, see Weinberg, 1995 (foundations) and Weinberg, 1996 (modern applications); an interpretive introduction to quantum field theory is given in Teller, 1997.
- 2 Einstein, 1916.
- 3 Lemaître, 1946.
- 4 Paty, 1993.

- 5 Einstein, 1949, in Schilpp 1949, pp. 683–684.
- 6 The basic reference about symmetry is Weyl, 1952; a pedagogical presentation of symmetry is given in Rosen, 1995.
- 7 *Réflexions sur la puissance motrice du feu et sur les machines propres à développer cette puissance.*
- 8 For a pedagogical discussion see Gell-Mann, 1994.
- 9 Maxwell, 1873.
- 10 For a discussion of the contributions of Gibbs and Einstein to the foundations of statistical physics, see Barberousse, 2002.
- 11 Gibbs, 1902, pp. xii–ix.
- 12 Einstein, 1917, p. 106.
- 13 Einstein, 1917, p. vi.
- 14 Einstein, 1917, p. 144.
- 15 Einstein, 1917, p. 156.
- 16 de Broglie, 1956.
- 17 Einstein, 1949, p. 668.
- 18 Einstein, 1935.
- 19 Weinberg, 1995, p. 177.
- 20 Weinberg, 1997.
- 21 Feynman, 1948, p. 370.
- 22 Gell-Mann, Hartle, 2006.
- 23 G. Cohen-Tannoudji, 1991, Lehoucq and Uzan, 2005.
- 24 Smolin, 2001, Padmanabhan, 2006.
- 25 Susskind, 2005.
- 26 Gonthier, 1967.
- 27 Lemaître, 1946.
- 28 Gonthier, 1949, p. IV-46 (310).

Bibliography

- Barberousse Anouk (2002), *La mécanique statistique. De Clausius à Gibbs*, Paris, Belin.
- Broglie, Louis de (1956), “Le dualisme des ondes et des corpuscules dans l’oeuvre d’Albert Einstein” (Discours à l’Académie des Sciences, Paris, 5 décembre 1955), in *Nouvelles perspectives en microphysique*, Paris, Albin Michel.
- Cohen-Tannoudji Gilles (1991), *Les Constantes Universelles*, Paris, Hachette.
- Einstein Albert (1916), *Die Grunlage der allgemeinen Relativitätstheorie, Annalen der Physik*, ser. 4, XLIX, 1916, 769-822; “The Foundation of the Generalized Theory of Relativity” in *The Principle of Relativity: Original Papers*, trans. M.N. Saha and S.N. Bose, Calcutta, University of Calcutta, 1920.
- Einstein Albert (1917), *Ueber die spezielle und die allgemeine Relativitätstheorie, Gemeinverständlich*, Braunschweig, Vieweg; *Relativity, The Special and the General Theory. A Popular Exposition*, New York, Crown Trade Paperbacks, 1961.
- Einstein Albert (1949), “Reply to criticism. Remarks Concerning the Essays Brought Together in this Cooperative Volume” in Schilpp, 1949.
- Einstein Albert, Podolsky B. and Rosen N. (1935), “Can Quantum-Mechanical Description of Physical Reality Be Considered Complete?”, *Physical Review*, vol. XLVII, pp. 777–780.
- Feynman Richard P. (1948), “Space-Time Approach to Non-Relativistic Quantum Mechanics”, *Review of Modern Physics*, Vol. 20, 2, p. 367–387.
- Gell-Mann Murray (1994), *The Quark and the Jaguar. Adventures in the Simple and the Complex*, New York, W. H. Freeman and Company.

- Gell-Mann Murray and Hartle James B. (2006), “Quasiclassical Coarse Graining and Thermodynamic Entropy”, (arXiv:quant-ph/0609190).
- Gibbs J. Willard (1902), *Elementary Principles in Statistical Mechanics. The Rational Foundation of Thermodynamics*, New York, Charles Scribner’s sons and London, Edward Arnold.
- Gonseth Ferdinand (1949), *La géométrie et le problème de l’espace*, Neuchâtel, Éditions du Griffon.
- Gonseth Ferdinand (1967), “Connaissance de la nature et connaissance philosophique chez Albert Einstein”, in *Science et synthèse*, Paris, Gallimard, pp. 21–28.
- Lemaître Georges (1946), *L’Hypothèse de l’Atome Primitif. Essai de Cosmogonie* (Préface de Ferdinand Gonseth), Neuchâtel, Éditions du Griffon.
- Maxwell James Clerk (1873), “Molecules”, *Nature*, Sept. 1873, pp. 437–441.
- Padmanabhan Tanu (2006), “Dark energy: Mystery of the Millenium” (Talk presented at the Albert Einstein Century International Conference at Palais de l’Unesco, Paris, July 2005) (arXiv:astro-ph/0603114).
- Paty Michel (1993), *Einstein Philosophe*, Paris, Presse Universitaire de France.
- Rosen Joe (1995), *Symmetry in Science. An Introduction to the General Theory*, New York, Springer.
- Schilpp Paul Arthur (1949), *Albert Einstein:Philosopher-Scientist*, The library of living philosophers, Lasalle, Ill., Open Court.
- Smolin Lee (2001), *Three Roads to Quantum Gravity*, New York, Basic Books.
- Susskind Leonard and Lindesay James (2005), *An Introduction to Black Holes, Information and the String Theory Revolution: The Holographic Universe*, Hackensack, N.J., World Scientific.
- Teller Paul (1997), *An Interpretive Introduction to Quantum Field Theory*, Princeton, N.J., Princeton University Press.
- Uzan Jean-Philippe and Lehoucq Roland (2005), *Les Constantes Fondamentales*, Paris, Belin, 2005.
- Weinberg Steven (1995), *The Quantum Theory of Fields*, vol. 1, *Foundations*, Cambridge, Cambridge University Press, 1995.
- Weinberg Steven (1996), *The Quantum Theory of Fields*, vol. 2, *Modern Applications*, Cambridge, Cambridge University Press, 1996.
- Weinberg Steven (1997), “What is Quantum Field Theory, and What did We Think It Is” (Talk presented at the conference “Historical and Philosophical Reflections on the Foundations of Quantum Field Theory,” Boston University, March 1996) (arXiv:hep-th/9702027).
- Weyl, Herman (1952), *Symmetry*, Princeton, Princeton University Press, 1952.

Foundations of Physics: The Empirical Blindness

Hervé Zwirn

1 Introduction

Physics has long been taken as the paradigm science. This was particularly the case under the logical empiricists. Physics was the only science that was worth discussing in epistemology. This is no more true and modern philosophy of science has to take all disciplines into account. Yet, it's true that biologists or sociologists don't wonder whether the objects they study are real. Their philosophy is a spontaneous realism which, in their mind, is not questionable. Physicists are the rare scientists wondering if scientific theories are about the world or about themselves. Physics remains the only empirical science that brings real new insights into philosophy and that is able to influence our philosophical conception of the world. Thus, the results of physics can't be ignored when discussing the status of reality or the validity of knowledge that science provides.

Among many others, these are questions that we can ask about physics:

1. Is physics describing the real world?
2. Is physics justifiable in any way?
3. Are the foundations of physics firm?

Of course these questions are linked. We'll try in the following to give some insight into possible answers.

2 Refusing Two Opposite Conceptions

Traditional epistemology pits realism against idealism. It is generally assumed that refusing the existence of an independent reality which has a precise structure and definite properties leads necessarily to a position close to idealism. I will try

H. Zwirn (✉)

Centre de Mathématiques et de leurs applications - UMR 8536

École Normale Supérieure de Cachan

61, avenue du Président Wilson

94235 Cachan cedex, France

E-mail: herve.zwirn@m4x.org

in the following to show that the choice between realism and idealism is not compulsory. Both positions could be wrong or more precisely only partially true. Both stem from a philosophical framework that is perhaps too narrow. The main point is that the competition between realism and idealism is often materialized through some questions that are supposed to receive either a positive answer (and in this case, for example, realism wins) or a negative one (and in this case, idealism wins). This dualism generally comes from the fact that the negation of a proposition is thought to be obtained by asserting the main verb of the proposition in its negative form: “It is false that A has the property P” is understood as equivalent to “A has not the property P”, which is sometimes assumed to imply that “A has the property not-P”. However, modern physics shows that sometimes, this is not the case. If it is false that the spin along Oz of an electron is $+1/2$, that doesn’t mean that this spin is $-1/2$, even if $+1/2$ and $-1/2$ are the only possibilities that could be obtained. When this sort of underdetermination arises for questions intended to decide between realism and idealism, neither the former nor the latter wins and the situation is more complex but also more interesting. I also want to focus attention on the necessity to avoid in this debate sentences like “something is real if one is compelled to somehow admit that it is different from nothing”.¹ The vagueness of this statement makes it both apparently true and in fact meaningless. But consequently, it can be used in many contexts, for example as an argument against idealism: “if thinking were not thinking of reality (understood as [...] something which is different from nothing), it would be thinking of nothing and therefore no thinking at all”.² This type of argument clearly doesn’t prove anything, but the dispute between realists and idealists often uses that sort of fuzzy (though apparently obvious) statement. Usually, the more these statements are obvious the more they are empty.

In the following I present arguments for and against realism and idealism and discuss them.

2.1 Scientific Realism

The scientific realist thinks that there exists an independent reality in which we are immersed and that this reality is literally and correctly described by scientific theories. Actually, scientific realism is made up of three assumptions. The first one is the thesis of metaphysical realism which claims that there exists an independent reality. The second one is the assumption that we can obtain some reliable knowledge of it. The third one states that scientific theories provide us with this knowledge.

Let’s clarify some of these points. Metaphysical realism says that reality is independent because it would be essentially the same even if we were not existing, in particular it is a domain of mind-independent existence. Of course, it is perfectly possible to accept metaphysical realism and to deny the fact that independent reality is knowable. According to Kant, we can’t have any knowledge of

things as they are in themselves. But this is not what the scientific realist believes. For him, we can know something about independent reality and the most plausible way to get this knowledge is through natural science. Thus the scientific realism thesis goes well beyond metaphysical realism. According to it, scientific theories give an appropriate account of the features of what objectively exists in the independent reality. That means that the scientific realist is entitled to believe that the entities whose behaviour is described by science are real in much the same way as a chair or a bird are real and that they behave as the theory says they behave. Accepting a theory is therefore accepting the existence of its objects. Van Fraassen³ says that according to scientific realism, science aims at providing us, through our theories, with a literally true story of what the world is. Thus, scientific theories are not to be thought of as metaphors but as expressing truths about the world. We must accept to the letter what they say. “To have good reason to accept a theory is to have good reason to believe that the entities it postulates are real” as Wilfrid Sellars has expressed it.⁴ If a theory is about electrons and their behaviour then the theory says that electrons exist. As Rescher⁵ puts it amusingly: “to accept a scientific theory about little green men on Mars is ipso facto to accept little green men on Mars”. So, physics describes reality such as it is in itself. According to the most recent theories, the string theories,⁶ we live inside a ten-dimensional space-time. Six of these dimensions are curled up very tightly so we may never be aware of their existence. Moreover the various particle types are replaced by a single fundamental building block, a “string” which can be closed or open and can vibrate. Everything in the world is ultimately made with strings. If we take scientific realism seriously then we must believe that strings really exist in a ten-dimensional space-time exactly in the same way we think that chairs exist in our ordinary space. It is even worse than that because what physics actually shows is that the usual objects we are used to are not really existing. That’s only the entities used in the theory that are existing. If we accept string theories then only strings exist. As Putnam⁶ says realism reminds him of the seducer in the old-fashioned melodrama. “The seducer always promised various things to the innocent maiden which he failed to deliver when the time came”. The maiden here is common sense, which believes that chairs and ice cubes exist and which is frightened by idealism (and all the similar anti realist positions). So, common sense naturally goes with the realist. Then after a while, the realist reveals to the poor common sense that it is not the chairs and the ice cubes that exist but the objects that scientific theories use, no matter how far from usual experience they can be. Putnam conclusion is “Some will say that the lady has been had”.

2.2 *Idealism*

For an idealist, the realist’s claim that things exist independently of our thought is inconsistent since to say something, it’s necessary to think about it. Thus simply by claiming that things are outside of our mind, these things are included in it.

According to Berkeley “*esse est percipi*”, only our perceptions are real. For him, the phenomena of sensations can all be explained without presupposing the reality of external material substances. Sensible objects exist only in the minds of those who perceive them. In its most extreme form, idealism leads to solipsism. The solipsist claims that only his thoughts exist. This position is not refutable but has the disadvantage to close the discussion. It was not Berkeley’s position. For him, what is real exists in many minds, so it can continue to exist whether I perceive it or not because somebody else can see it. But the problem is to account for the fact that the objects that I perceive now continue existing even when neither myself nor nobody else perceives them. For Berkeley, God plays that role. The mind of God serves as a permanent repository of the sensible objects that we perceive at some times and not at others. For Hegel and the German idealists this role was played by the Spirit, self-knowing, self-actualizing totality of all that is, obtained through dialectical reasoning as the synthesis of the thesis Idea and its antithesis Nature.

3 Arguments for Realism

3.1 *Metaphysical Realism*

Metaphysical realism seems a natural position directly based on the spontaneous knowledge of the world we acquire since early childhood. Believing in an external world in which we live and that doesn’t depend on what we think or know is a natural attitude directly drawn from our everyday life. As Hume noticed, in general the reason why we think that something exists is based on the “cause-effect relation”. If I hear a voice in the room next door, I think that there is somebody speaking. If I see a shape looking like a plateau with four legs, I think that there is a table in front of me. We assume that something exists that causes the effect we see and indeed, our experience shows that in many cases, this reasoning is correct (we’ll see below how Hume refutes the validity of this argument). Another argument for metaphysical realism is the fact that people agree on what they see. Two persons watching a garden will agree saying that there are two trees, a clump of flowers and a dog on the grass. In general, people’s observational reports are in agreement and why should it be so if the observations were not about something which really exists independently of the persons who observe? We’ll call this argument the *inter subjective agreement*. Another reason for believing that something that doesn’t depend on our theories or on our thoughts exists is the fact that we can’t do what we want. Of course, it is clear that we can’t fly but more than that, our scientific theories sometimes fail and are refuted by experiment. If there exists a reality that has a proper structure, we can’t say anything and everything about it. There are propositions which will turn out to be false as they state what it is not.⁸ D’Espagnat⁹ says that “there is something that says no”. The fact that there are propositions in the scientific theories which turn out to be refuted through an

adequate experimental device is an important clue that these propositions speak of reality. Thus, the three main reasons for metaphysical realism are the cause-effect relation, the inter subjective agreement and the resistance of reality (something that says no).

The most elementary version of metaphysical realism is the layman's spontaneous philosophy: independent reality is made up of objects such as chairs or cubes of ice (and perhaps also of waves and of forces for the most advanced ones). In this conception that d'Espagnat¹⁰ calls "multitudinism", the world is nothing more than a collection of entities with well defined properties, that exist independently of us and interact in a well defined way. Of course, much more sophisticated versions of metaphysical realism are possible. For example, Kant's transcendental idealism is compatible with a sophisticated metaphysical realism which asserts that we can't have any direct access to things as they are in themselves (noumena) and that the only access is through our experience (phenomena).

3.2 Epistemic Realism

The very same reasons that make us believe that there exists an independent reality also lead us to think that we can know something about it. After all, we have learnt since the beginning of our life to use the objects around us, we know what is going to happen if we let a ball fall down, we wait for the daylight after the night and we drink water when we are thirsty. Thus, we know something, we even know many things. Common sense tells us that what we know is about reality. So reality is knowable. There is no reason to think that we can't know anything about reality since we do know a lot of things about it. Now, common sense and everyday life also tell us that many phenomena that we see are easy neither to understand nor to forecast. When will the next eclipse come? Why is the sky blue? How can we cure flu? These questions are not easily answered and that is where science becomes necessary. So science, being nothing else than the continuation of common sense by other ways as Bertrand Russell says, inherits the quality of common sense to speak about reality.

3.3 Scientific Realism

It is then natural to adopt the position according to which the best way to get knowledge about reality is to ask science for it. It is difficult to deny that almost all scientists adopt a spontaneous realism. According to most of them (and this is all the more true as we move away from theoretical physics to biology and human sciences), their work aims at describing reality (or at least some part of it) and they believe that science can succeed in this objective. The most usual argument to defend scientific realism starts with the acknowledgement that scientific

theories succeed in giving a correct description of the observed phenomena and in making it possible to forecast them at least within good approximation, what Boyd¹¹ calls the instrumental reliability of scientific theories. This success would be very surprising if science was not describing what reality is in itself. This is the “*no miracles*” argument of Putnam and the “*abductive argument*” of Boyd. This argument states that it is only by accepting the reality of approximate theoretical knowledge that we can adequately explain the uncontested instrumental reliability of scientific methods. The fact that there is an independent reality and that it is described by scientific theories is the best explanation of the fact that science is working. Put differently, this means that the explanation for the empirical success of a theory is simply that this theory is true. And if the theory is true then the entities that it deals with are real, exist and behave as the theory says they behave. An important consequence of this position is that we are entitled to believe also in the non-observable entities of the theory. For example, even if quantum chromodynamics tells us that it will never be possible to observe a free quark directly because of confinement,¹² quarks are really existing in much the same way as atoms are.

4 Criticism of the Arguments for Realism

4.1 *Metaphysical Realism*

4.1.1 Cause-Effect Relation

If I hear a voice in the room next door, I think that there is somebody speaking. If I see a shape looking like a plateau with four legs, I think that there is a table in front of me. The first sentence means that we infer from one perception (the voice in the room next door) another possible perception through a counterfactual reasoning: if I went in the room next door, I would see somebody. The second one concerns the inference of the existence of something from a perception. These two inferences are not of the same type. Criticisms against both have been raised a long time ago. The first inference only concerns perceptions, it links two perceptions between them. Each time I hear a voice in the room next door, if I go into the room I'll see somebody. Hume criticises this inference which rests on the principle of induction that states that if such a link has been observed in the past it will remain valid in the future. Now induction principle is impossible to rationally justify. Thus nothing guarantees that two events that were linked in the past will remain linked in the future. So, from a purely rational point of view, we are not entitled to think that we'll find somebody when we hear a voice in the room next door. Induction is not a valid inference. We know that the solution given by Kant to this problem is to consider that induction must be understood as a synthetic a priori condition for empirical discourse. The second inference concerns reality itself. It consists in assuming the existence of real entities as an explanation for

our perceptions. Thus only this second formulation concerns directly metaphysical realism. Hume criticized it noticing that we have access to our perceptions only and not to reality. That things exist outside is thus not a conclusion that we should draw. According to him, it is only as a convenient way to organize our perceptions that we are led to assume that external objects exist. It is true that from a logical point of view we have no valid reason to fill the conceptual gap between the existence of our perceptions and the existence of an external world. After all, the objects that I perceive when I am dreaming are not external to my mind. Moreover, quantum mechanics gives good reasons to be prudent about conclusions such as “this thing exists”. Thus, I will not consider the cause-effect relation as a strong argument for metaphysical realism.

4.1.2 Inter Subjective Agreement

Inter-subjectivity is an argument for the existence of things external to our mind based on the remark that pure idealism which states that there is nothing outside, has difficulty to explain why we agree about our perceptions. If Paul and Peter both agree on the fact that there are two glasses and a bottle of wine on the table, the simplest explanation is that there are really two glasses and a bottle of wine on the table. In some sense, this argument is an answer to the objection I raised above to the cause-effect argument: if I perceive a table, that could be an illusion (as in a dream). The table could exist only for me. On the other hand, if both Paul and Peter see the table, it becomes hard to say that they share the same illusion or, even if this is possible sometimes, it seems difficult to claim that all common perceptions are illusions. So, inter-subjectivity seems to be more a solid argument for metaphysical realism than cause-effect relation. And yet, quantum mechanics undermines this conclusion. This is not because Paul and Peter agree on the fact that the spin along Oz of one electron is $+1/2$ that this spin was equal to $+1/2$ before the measure. The intuitive explanation saying that if they both see that the result of the spin measurement is $+1/2$, this is because the spin was $+1/2$ before the measurement, is wrong and leads to false consequences. Quantum mechanics teaches us that the value of the spin was indefinite before the measurement and that it has become determined during the measurement process. In some sense, it is Paul and Peter’s perception that is (partly) the cause of determination of this value. Quantum mechanics says that the very fact of measuring is, at least partly, the cause of what we perceive. This is particularly clear in the interpretation of the measurement process given through the decoherence theory.¹³ Now, even if we contest the conclusion that if Paul and Peter have the same perception of a table this is an argument for the existence of this table, we have admitted till now that their perceptions were actually identical. But this is not mandatory and within the position that I have called “convivial solipsism”¹⁴ I propose an interpretation of the measurement process where inter-subjectivity is apparently respected though perceptions are different. So, inter-subjectivity seems at the end not to be as strong an argument as it could appear intuitively.

4.1.3 Resistance of Reality

If reality was only a human construction there would be no reason why our best theories be refuted by experiment. But the history of science attests that there has been an uninterrupted series of refutations. A lot of beautiful and powerful theories have been defeated by experimental results that were not in agreement with their predictions. So, as d'Espagnat puts it, "there is something that says no". And according to him, this "something" can't be "us". This argument implicitly assumes that a human construction will be its own yardstick and will be exempt of contradiction. But why should it be so? If we suppose that everything comes from us, that we invent the rules of the game, what we identify with reality is a human construction. Of course, it is very difficult to explicitly describe the nature of this construction and the way we use to build it. It is totally different from the formal way we build scientific theories. We start building it in our early childhood. Let's call it a perceptual construction. Scientific theories are then explicit, conscious and formal constructions intended to account for an unconscious perceptual construction. We know how difficult it is to show that a formal system is consistent as soon as it is complex enough. Gödel's theorem proves that we can't demonstrate the consistency of a formal system by purely internal means (as soon as the system is powerful enough to contain arithmetic). So, building a consistent complex system is not an easy task, the more complex the more difficult. By extension¹⁵ it is not so surprising that from time to time we discover some contradictions, which manifest themselves through empirical discrepancies between our formal scientific constructions and the informal construction that we call reality. That means that we have some difficulties to build two different constructions (one formal giving scientific theories and one perceptual giving what we call reality) in such a way that these two constructions be simultaneously consistent. D'Espagnat¹⁶ answers that, if everything is nothing but a construction coming from us, he doesn't understand why we generally choose to preserve the construction that represents reality against the theoretical construction. He says that we could as well choose to believe in a refuted theory and abandon our belief in reality. The reason why we don't do that is because the two constructions are not on an equal footing. Reality is much more epistemically entrenched than science, and confronted with the necessity of revising our beliefs we always choose to change those that are the less entrenched.¹⁷ Science is by definition an empirical process and that means that confronted with a contradiction between data and theory we must give the priority to data.

4.2 *Scientific Realism*

4.2.1 **Abductive Argument of Empirical Success: No Miracles**

As we have seen the abductive argument for scientific realism is often considered as the main argument for it. The empirical successes we get in applying our theories betoken their truth. This argument is also indirectly an argument for metaphysical

realism since scientific realism rests on metaphysical realism and that assuming the truth of a theory without assuming a reality to refer to would be meaningless. Oversimplified this argument is nothing else than the old explanation that we see the grass green because grass is really green. It is well known since Locke's distinction between primary and secondary qualities that this is an explanation which raises many problems. Moreover, there has been an uninterrupted series of refutations of momentarily adequate theories during the history of science. That shows that it is difficult to believe that our current theories, even the best ones, are true because using a pessimistic inductive argument, we are led to think that they will be refuted in the future. The only possibility to save this argument is then to adopt (with Boyd¹⁸ for instance) the concept of a gradual convergence of scientific progress. We'll show below that this is not acceptable. Another criticism can be given along the following line. Given a finite set of data (resulting from observations) it is in principle possible to build many theories accounting for them (as there is an infinity of curves going through a finite set of points). This is Quine's thesis of underdetermination of theory by evidence. So, there is no reason to be surprised that we can build adequate theories at a given time. The defenders of the argument admit this point and retort that it is not the description of known facts but the prediction of novel facts that would be miraculous if the theory were not reflecting something real. The most often quoted example is the discovery of Neptune through pure computation within Newton's mechanics. But, let's be cautious not to fall into the illusion of what can be called the horoscope effect which is the fact to remember only the successes and to forget the failures. The discovery of Neptune is of course a very remarkable prediction but the inexistence of Vulcan whose mass and position had been calculated to explain the correct value of the precession of the perihelion of Mercury is a memorable failure too! As Popper says, science is going on through conjectures and refutations. At each period of time, the stock of empirical data is finite. So it is in principle possible to account for it through many theories. Sometimes, scientists provide competing theories that need to be tested to know which is the best one. Some of these theories predict novel facts (a new planet or a new particle or a new physical effect). The best corroborated theory is kept but we must have in mind that other competitors, predicting other novel facts, have been refuted. To give a recent example, several theories were in competition with the Glashow–Weinberg–Salam model (now called "the standard model") to unify weak and electromagnetic interactions before the discovery in 1983 of the W and Z bosons predicted by the theory. Should these bosons have been inexistent, another theory would have survived. *A posteriori*, it seems a miracle that the winner predicted this novel fact but is it really surprising? When many theories compete, it is not strange that sometimes one of them be momentarily correct.

Besides that, one can wonder if it is really necessary for a theory to be true to provide good results. After all, nature could be error-tolerant and if the error has an impact that is below the threshold of the accuracy of today's observations, then the theory will be confirmed. As Rescher¹⁹ says:

The success of the applications of our current science does not betoken its unqualified truth or ultimate adequacy. All it indicates is that those various ways (whatever they may be) in

which it doubtless fails to be true are not damaging the achievement of these good results – that, in the context of those particular applications that are presently at issue, its error lie beneath the penalty level of actual failure.

So the empirical success argument doesn't seem really convincing. It is not necessary to assume any correspondence between a theory and reality to understand its empirical success. We'll see another reason not to accept this argument when we speak of scientific theories as compression algorithms.

4.2.2 Non Convergence of Scientific Theories

In this view, scientific theories are becoming closer and closer to the truth and converge gradually towards an ultimate (perhaps forever out of reach) true theory. Though this conception may seem appealing it comes up against many difficulties. The first one is that we know since Popper's unsuccessful definition of verisimilitude that we have no satisfactory definition of approximate truth for a theory. What does it mean that a theory (which is known to be false) is closer to the truth than another one? In which respect is Newton's mechanics (which has been proved false after the discovery of special relativity) closer to the truth than the Ptolemaic theory? There is one meaning of approximation which is unproblematic. This meaning is related to the numerical predictions made by the theory. If these numerical predictions are in general more accurate within one theory than within the other, we are entitled to say that the first one is numerically closer to the truth than the second one. And this is exactly what happened during the history of science. A theory empirically adequate at one time was replaced by a new one because its predictions were either in disagreement with experiment (that appeared through a numerical discrepancy) or less accurate than those given by the new one. Let's give as an example the successive replacement of the Ptolemaic system by Copernicus' circular trajectories then by Kepler's laws then by Newton's mechanics. At each step the predictions have been improved, which was necessary due to the improvement of observational means and of the accuracy of measures. Though empirically adequate at the time of the Greeks, the Ptolemaic system is refuted by modern results of observation. Today we know that Newtonian mechanics (empirically adequate long after Newton's time) is refuted too. Its predictions are less and less good as speeds closer and closer to the speed of light are considered, and special relativity is the currently best theory in this case. Newton's theory is also refuted by the precession of the perihelion of Mercury, which is only accurately predicted by general relativity. As computed inside Newton's theory, there is a discrepancy of 43 s of arc less per century. So, it is meaningful to say that as far as numerical predictions are concerned, Einstein's relativity is closer to the truth than Newton's mechanics, which was closer to the truth than Kepler's laws, which in turn were closer to the truth than the Ptolemaic system.

But this is not what realists have in mind when they say that a theory is closer to the truth than another one. They go further, meaning that what the better theory says

about reality (the objects that the theory describes, the laws it uses, the structure of reality which is implicit in its mathematical structure) is reflecting more truly what is real. This is where we can't follow them. Literally speaking, the Ptolemaic system and Newton's theory are false. From the point of view of Einstein's relativity, there are no epicycles, there is no absolute time and there is no gravitational force acting at a distance. So, what could it mean that Newton's theory is closer to the truth than the Ptolemaic system apart from the accuracy of numerical predictions? If the assumed truth is the description given by general relativity (which is the best theory for gravitation we have today and so, in this view, is supposed to be the closest description of the truth we ever had), is the image given by a flat space with an absolute time and a gravitational force closer to the truth than the image of epicycles? It is highly dubious. Actually, if we analyse the realist's reasoning, it is very loose. It uses one fact and one hypothesis. The fact is that there is no doubt that the empirical adequacy of scientific theories is increasing as far as the accuracy of the predictions is concerned. The hypothesis is that an empirically adequate theory must be close to the truth (unless it is a miracle that it can give correct predictions). Truth means here that the structure of the theory closely reflects in every aspect the structure of reality and that the entities of the theory refer to real objects. Then from the fact that theories become closer and closer to numerical truth they infer that theories become closer and closer to the truth (truth taken in the above sense). This is clearly a wrong inference since it demands another assumption to be valid: the assumption according to which, having defined the concept of distance to the truth of a theory (let's call it f , so $f(T)$ is the distance to the truth of the theory T), this function has the following property: if T' is numerically better than T then $f(T') < f(T)$. There are clearly two problems. The first one is that we don't know how to define this function f , the second one is that, even if this growth property of f seems intuitive, it is not obvious that it is not possible to define a distance to the truth violating it.²⁰ To summarize the argument: the empirical successes of a theory is supposed to betoken its truth. Faced with the objection that it can't be the case since many successful theories have been refuted and so were false, realists answer that these theories were not totally true but close to the truth. Thus, the argument has moved from "a successful theory is true" to "a successful theory must be close to the truth". But for lack of the definition for approximate truth, the only possibility for realists is to notice that theories are becoming more and more accurate and, under the implicit assumption that the more accurate the closer to the truth, to infer that theories are converging towards the ultimate true theory. This is not convincing since they are unable to define what they mean by close to the truth if this is not in the numerical sense.

There is also another reason why this argument is problematic. It is based on the hypothesis that there is an ideal (though perhaps out of reach) theory which is true. This theory is totally in agreement with reality, which is supposed to be correctly described in all respects by it. But because of Quine's underdetermination of theory by evidence there is not a unique ultimate theory but presumably many that are empirically adequate whatever the stock of empirical data be. These theories could be incompatible or even contradictory in many ways. The famous example given by

Putnam²¹ is the one of two theories T and T' which are empirically equivalent but such as T entails that "there really are such things as spatial points" and T' entails that "there are arbitrarily small finite regions but not points". In this case, which one is representing the truth? How is it possible for two contradictory theories to be true? Are there many truths? One hits a paradox that seems not easy to escape. Putnam's internal realism enables us to say that both theories are true. But this is clearly because Putnam declines to assert a theory of truth and in particular denies that truth is captured by correct assertability.²²For a traditional scientific realist this is clearly a problem.

4.3 What is Left in Defence of Realism?

It is true that common sense exerts a very strong influence on the feeling that we live in an independent reality that frequently resists against what we want to do or that refutes what we could think a priori about things around us. It is also true that it appears that we do know something about reality, first through our everyday experience, second (in a more precise and efficient way) through our scientific theories. The proof is everywhere around us. We have invented planes which allow us to fly, rockets to go to the moon, satellites that allow us to communicate everywhere on the earth and even to find our location everywhere with a precision better than one meter, we are able to extract energy from atoms, etc. This list could be made much longer but this would not add anything to the argument. The force of what we see, what we hear, what we feel, what we experiment is extremely strong and leads us to become unable to envisage the possibility that there is no independent reality. Our feeling about its existence is shaped day by day since our childhood and is strengthened as we get older through the knowledge that we get from science. The process is the same for the child and the scientist. The abductive argument plays its role, and it is a very pervasive role! But, remember: we have this very strong feeling that we are at rest and yet the Earth is moving around the Sun at a speed of 28 km/s; we think that a particle must have a definite position and quantum mechanics teaches us that this is not always the case; we feel that time goes everywhere the same way, but relativity theory shows that it is false; we believe that energy is a property of objects, and yet relativity theory shows that it is possible to transform energy into particles. All these strong feelings are extremely entrenched in our mind. Abandoning them is a very difficult task that many people refuse to accomplish (physicists excepted). Though these feelings mislead us. Can we think that our feeling that there is an independent reality is similarly a misleading feeling? We have given above good reasons for that and will propose in the following a conception that doesn't need this hypothesis.

5 Arguments for and Against Idealism

I will be much shorter as far as idealism is concerned. Berkeley's idealism was perfect for somebody wanting to defend religion and God against the dangers of science, materialism and atheism. His claim is that sensible objects cannot exist without being perceived. It is true that resorting to God for allowing things to exist even when no human being is perceiving them is both a means to answer the main criticism against idealism – it is very difficult to accept that this tree that I am the only one to see in the deep forest vanishes every time I close my eyes to reappear when I open them again – and a good argument in favour of the existence of God. The intellectual contortions of the German idealists to avoid to resort to God are not very satisfactory. The conception I will present below, although not realist, avoids these problems because it doesn't assume that everything comes from our mind.

I will now present some ingredient of my conception.

6 Scientific Theories as Compression Algorithms

6.1 *The Algorithmic Theory of Information*

The algorithmic theory of information has been simultaneously invented by Kolmogorov, Solomonoff and Chaitin. The algorithmic complexity of a string of bits is the smallest self-delimiting program to produce it. This measure is essentially concerned with the redundancy inside the string. If there is no redundancy, no regularity then the only way to produce the string is to give explicitly the string to the computer. We agree that the string "010101010101010101" is simple because it can be described as a regular succession of "01". The string "0100110000110000011110010101" seems on the contrary more complex because it is difficult to discern any pattern inside. A regular, ordered string will be made up of 0 and 1 that follow a simple rule. A string of one thousand "01" is easy to produce through a small program. This program is simply "write one thousand times "01" and stop". The important fact to notice is that in this case, it is possible to produce the string through a program that is significantly shorter than the string it produces. Typically the size of the program is roughly the logarithm of the size of the string. On the other hand, a string as "0100110000110000011110010101" will not be computed through a program using a simple rule, just because there is no simple rule to do the job. The shortest program to produce it will then be "write "0100110000110000011110010101" and stop". And the size of this program is roughly the same as the size of the string. The main lesson to remember

is that when a string is ordered, when the succession of 0 and 1 follows a simple law, then it is possible to compute it through a short program, a program that is significantly smaller than the size of the string. On the other hand, when the string doesn't contain any regularity, when the succession is random then the only way to compute it is to give it explicitly to the program because no rule can compute it. Then the size of the program is the same as the size of the string.²³

6.2 *Scientific Theories as Algorithms*

Let's now view scientific theories as formal systems. Given one theory, the set of experimental data belonging to its domain is a set of statements inside the language of the theory. Now the theory will be empirically adequate if it can produce these statements, i.e. if these statements are provable inside the theory.²⁴ So the theory, if adequate, is a way to compute the empirical results. Thus, it becomes natural to see the theory as an algorithm to produce the set of empirical results. Let's take as an example, the series of experimental results from successive releases of a pebble in the vacuum starting from a height of one meter to a height of one hundred meters through one meter increments. The speed of the pebble when it lands is a one hundred numbers series. This series can be produced from a very simple algorithm given in Newtonian gravitation theory: the formula $v = (2gh)^{1/2}$. This simple example can be generalised to all predictions done by physical theories, even the most complex ones. Of course, all physical phenomena don't reduce to a series of numbers. Finding the good algorithm often needs to find good tools or new concepts to describe the phenomena that are studied. This is why, building a physical theory doesn't reduce to finding mathematical formulas. But, this doesn't change anything to the fact that, once the theory is built, it is an algorithm making it possible to generate the statements reporting experimental results. An empirically adequate theory must of course be able to produce all the experimental data coming from past experiments but it attempts also to predict the potentially infinite set of results of every future experiment. In this sense, the theory is really a way to compress the infinite number of data coming from all potential experiments that could be done inside a physical domain. As Chaitin²⁵ says:

I think of a scientific theory as a binary computer program for calculating the observations which are also written in binary. And you have a law of nature if there is compression, if the experimental data is compressed into a computer program that has a smaller number of bits than are in the data it explains. The greater the degree of compression, the better the law, the more you understand the data. But if the experimental data cannot be compressed, if the smallest program for calculating it is just as large as it is [...] then the data is lawless, unstructured, patternless, not amenable to scientific study, incomprehensible.

From this point of view, comprehension is compression. It is then possible to say that the domain of science is the set of all phenomena that are compressible. That's not the case for art for example. It is not possible to find an algorithm decid-

ing if such a painting is better than another one from the aesthetic point of view. It is not possible to build an algorithm to produce grandiose symphonies. So there is no more reason to be surprised with Einstein that the world (the physical one) be understandable or to find extraordinary with Wigner that mathematics be so efficient. The reason why is that we apply science only where it is possible, that means in the domain of compressible phenomena. The only surprising thing is then that there are such phenomena!

6.3 *Induction Revisited*

Hume's devastating criticism against induction shows indubitably that induction is not a valid reasoning. However, and contrary to Popper's claim, it is commonly used in scientific reasoning. In the light of scientific theories seen as compression algorithms the reason why it is so becomes clearer. Scientific process starts by building theories that explain (reproduce) past empirical results and then, use them to predict new results. If we think that scientific theories aim at compressing data, the best theories will make the best usage of the regularities that data contains. As we have seen, algorithmic complexity addresses the regularity that a string of 0 and 1 (or more generally a series of results) contains. If there is any regularity, the minimal program (program of the shortest length that computes the data) must certainly exploit it. For example, assume that the data you have gotten from your past experiments is a string of one million 1, representing the results of one million experiments. The minimal program to produce this string is certainly "write 1 one million times". From our point of view of scientific theories as compression algorithms, this is not exactly what we would like to call a theory. The corresponding theory would be an algorithm giving you an answer when asked a question. In this (too) simple case, the question is "what is the result I'll get if I do an experiment" and the answer is "1" which is confirmed in every past experiment. Now, if we want to use this program for predicting new data, the simplest way to do it is to reproduce the same answer. The best theory under this point of view will be "all experiments give a result 1". Faced with a series of observations sharing common features (every morning the sun rises), given the fact that the smallest program that predicts this data is exploiting this regularity, the best theory will be the one that continues predicting the same features under the same conditions. So, the usual reason why we use induction, that is the belief that things that happened in the past will repeat identically in the future is replaced by the precept that, having built an algorithm which reproduces past experiments through the shortest way, we want to use it to build the simplest and shortest algorithm to predict the future. It happens simply that induction, i.e. reproducing past regularities, seems the right way to do that. Thus, that means no more that we must have confidence in the fact that future is similar to past, that means that following this particular rule for building scientific theories, we must believe that (1) minimal programs are the best theories and (2) using regularities is the best procedure to build minimal programs. So the induction justification shifts from a metaphysical problem to a pragmatic one.

7 The Empirical Blindness

7.1 *Empirical Pertinence*

As is well known, the positivist demarcation between observational and theoretical statements has been questioned first by Hanson²⁶ then by Kuhn²⁷ and Feyerabend.²⁸ Observations and experiences have to be interpreted to be meaningful and that involves an inescapable theoretical dimension. That is what is called theory-ladenness. One consequence is that despite its harmless appearance, the fact to accept a theory as empirically adequate is less neutral than one could think. Accepting to recognize that a theory T has been, up to now, empirically adequate is already showing a certain commitment to T. As van Fraassen says,²⁹ it is in particular accepting to use the conceptual framework of T to guide the look for new experiments and to interpret the observations. It is also accepting the fact that the set of observations made to test T, which have been guided by the research program induced by T, is relevant and significant compared to all the observations that could have been made to test T. T will be considered as empirically pertinent if the research program induced by T is relevant to guide the experiments made to test T. For example, a theory that would predict that every time there are clouds in the sky, a new moon will come within less than 28 days would be empirically adequate but not empirically pertinent. Accepting a theory as empirically adequate requires a preliminary commitment which is justified only if first, the structure of T is not too far from the dominant paradigm on good theories and second, T has some successes in its favour such as an explanation of a fact not understood in previous theories or the prediction of a novel fact.³⁰ There exists then a mutual support: the empirical successes got in the framework defined by the theory increase our confidence in the pertinence of the theory which in return confirms us in the fact that these successes are good evidence for the empirical pertinence of the theory. This is similar to the process described by Boyd³¹ “there is a dialectical relationship between current theory and the methodology for its improvement”. But unlike Boyd who interprets this as a possibility for a cumulative development of science, I think on the contrary, that it is a flaw weakening the status of empirical pertinence. Why? Because this process, once engaged in a false direction can maintain wrongly its own success.

7.2 *Empirical Blindness: First Aspect*

This mistaken process can happen through many different ways. The first one consists in providing predictions fuzzy enough for being confirmed whatever the experiments or in remembering only the observations that confirms the predictions. The best example of this kind of practice is astrology. Predicting to somebody who is a Libra that something concerning his family is going to happen during the week is not a very risky prediction since at any time almost everybody can refer to a familial

event happened recently. Moreover, when asked about the quality of predictions made to him, a staunch supporter of astrology will remember only the successes and forget the failures. A second one consists in using ad hoc modifications to escape falsification. The famous example given by Popper is Marxism because its advocates preserved the theory from falsification by modifying it and because the only rationale for the modifications which were made to the original theory was to ensure that it evaded falsification. These two ways to ensure the success of a theory pose the problem to know how distinguishing a good scientific methodology. No definitive precise demarcation criterion is accepted yet. We have acquired a good intuition of what science is but this intuition is difficult to formalize. Nevertheless, we can think that this intuition is good enough to prevent us from falling into the traps just described. Yet, there is a more subtle possibility to maintain a wrong empirical success. This will be the case if the dominant theory provides a conceptual framework such that no experiment that could falsify it be launched. Here, Lubbock's famous quotation "what we see depends mostly on what we look for" is particularly appropriate. Let's use a voluntarily exaggerated example to show what we mean. Imagine as Putnam that there is a Twin Earth where the laws of nature are different. Let's call it Earth 2. On Earth 2, there is an absolute space and a universal time. There is no need neither for special nor for general relativity. The Newtonian mechanics (its local equivalent invented by Newton 2) is the dominant theory T and it works perfectly well. Roughly, Earth 2 is at the same scientific level than we were at Laplace's time. The main difference is that under the hypothesis we have adopted, no experiment (such as a discrepancy in the precession of the perihelion of Mercury) can falsify T predictions for macroscopic bodies. Assume moreover that the paradigmatical impact of T is such that all researches are focused on the behaviour of macroscopic bodies and that scientists are blind to other phenomena. On Earth 2, T works perfectly well to describe the behaviour of macroscopic bodies within the accuracy of the most precise experiments. The nature of light is not considered as a scientific problem and neither electricity nor magnetism have been discovered. In this case, T is empirically adequate since it is used only in a restricted domain where it works and only experiments about this domain are considered scientific. For Earth 2 realist physicists, T is true and reality is made up of objects like stones or planets following T laws. In this case, the empirical adequation of T comes from a bad empirical pertinence. The program of research induced by T has led physicists to be blind preventing them from launching experiments needed to refute T as for example making light waves interfere or study black body radiation. Of course, it is easy to raise objections against this too simple example. One could say that in a world where there is no need for relativity, the laws of Nature could well be as Earth 2 physicists think they are. I ask the reader to accept as a hypothesis the fact that it is possible that on Earth 2 everything is as it is on our Earth except for relativistic behaviours. The second, more serious, objection could be to criticize me on the basis that I have assumed what I wanted to prove in postulating that Earth 2 physicists don't care about the nature of light and that neither electricity nor magnetism have been discovered. My answer is that my example is certainly too simple but that I use it to show that the program induced by a theory can lead to forget

about certain not obvious phenomena. Sometimes, having the idea to test something requires a long and difficult preliminary theoretical work. Aspect's experiments that showed that locality is violated have been possible only after Bell's discovery of his inequalities. The non-separable property of space that follows comes from a very sophisticated experimental device built on a complex theoretical proof. This is not something that is just in front of our eyes. One proof is that neither Einstein nor Bohr during their hard discussions about EPR paradox have been able to imagine a real experiment to decide between their positions. So it is not absurd to think that for some theories, such complex phenomena, that could refute the theory if they were seen, stay hidden inside the framework of the theory. In this case, the theory, though false, will never be falsified. Something similar happened (admittedly during a short period of time) with mechanics when physicists at the beginning of the 20th century were exclusively focused on integrable systems whereas we know now that non-integrable systems are the vast majority. Empirical blindness is then the fact to be blinded by a research program coming from a limited dominant theory that hinders from doing the experiments that could refute it. As this theory is not empirically pertinent, some phenomena remain then ignored and even can be left outside of scientific preoccupation.

7.3 Empirical Blindness: An Inevitable Disease

There is another aspect of empirical blindness which is linked to undecidability. As is well known since Gödel, any complex enough formal system contains propositions that cannot be proved nor refuted inside it. These propositions are called undecidable. The consequence is that if some phenomena are expressed through such undecidable forms inside a theory, the theory will have nothing to say about them. Let's give an example. ZF is the Zermelo Fraenkel set theory. The totality of mathematics used in physics is inside ZF. Let's call T, a physical theory built by adding to ZF, what is needed to get a good physical theory as for instance the M-theory we mentioned previously. Let's call ZFE, the theory ZF to which we add some large cardinals hypotheses.³² The consistency of ZFE can't be proved inside itself and even less inside ZF and for some large cardinals hypotheses nobody has a reliable feeling about the fact that it is consistent or not. Now, it is perfectly possible to build a real Turing machine enumerating one by one all theorems of ZFE and stopping when it proves "1=2". Then, T will be dumb about the question whether the Turing machine will stop or not. This can seem strange but such a machine is a real physical system and no physical theory can predict its behaviour. This impossibility doesn't come from lack of knowledge of some laws or from quantum effects and would remain even in a classical world governed by pure mechanical laws. This is an example of a simple physical system whose behaviour is forever not predictable. The algorithmic power of scientific theories can't include the totality of empirical phenomena. This is another aspect of empirical blindness. Every theory will stay blind to some parts of the empirical reality and its formalism will stay dumb on the behaviour of these parts.

8 A Three Levels Stance

8.1 *The Realism of Phenomena*

Many positions (realist or not) assume that phenomena are explicitly given to us as if empirical reality was a big bag which we can get phenomena from or a stage where phenomena happen and can be looked at. This is what I call the realism of phenomena. It is, relative to phenomena, a position comparable to naïve realism which says that objects are really existing, are under our eyes and that it is the reason why we see them. This is a natural position for anybody accepting metaphysical realism. Oddly enough, it is sometimes admitted also by idealist thinkers who reject metaphysical realism. They refuse to accept the existence of objects in themselves but admit that we have a direct access to phenomena identified with our perceptions. For these thinkers, we content ourselves with watching passively what happens. This doesn't imply that we make no effort to observe phenomena but that these efforts are more like bending down to pick up a pebble than like creating something that would not exist otherwise.

Usually, two levels are considered. The first one is the level of phenomena (identified with our perceptions) and is called empirical reality. The second one, rejected by non-realists, is reality in itself. If realists admit that reality in itself exists beyond empirical reality, many non-realists admit that empirical reality exists in a sense that is very close. I don't agree with this position and I feel the need to introduce a distinction between what I'll call empirical reality (with a new meaning that I am going to precise) and the level of our perceptions that I call phenomenal reality. I will defend the following stance: our perceptions, interpreted, form the basis which we rely on. These perceptions are not to be considered as simple awareness of some pre-existing empirical reality. According to the realist image, phenomena exist outside, and we become aware of them through our perceptions. What I say is that no phenomenon exists outside and that we are responsible for our perceptions. In some way, we create phenomenal reality. But we are not free to create it as we want because many constraints are there. These constraints are what I call empirical reality. Empirical reality is what makes our perceptions possible while imposing constraints on them. That means that it is the framework for all the (physical and mental) acts that we use during the cognitive process. It is the set of all potentialities which, during their actualisation, give birth to our perceptions. Let's give an analogy from quantum mechanics. Perception and potentiality are in the same relationship as the result of a measurement (for example the value of the position of a particle) and the physical property that is measured (the position). It is through the measurement process that a value for position is determined (and that we become aware of it), but the position is not determined before and as such doesn't pre-exist as phenomenon but only as potentiality. Thus, a phenomenon is not something that pre-exists and that we observe passively, it is something that emerges from an act in which we play an essential role. Dummett³³ defends a similar position in mathematics and copying what he says and adapting it to physics,

I would say: If we think that our perceptions come from outside (and I call phenomenal reality the set of all perceptions) we should think of a phenomenal reality that is not existing yet but that comes to birth along with our actions. Our tentative research gives birth to something that was not there before and what it gives birth to (phenomenal reality) is not coming exclusively from ourselves. Empirical reality is the framework that constrains our perceptions and makes them come from outside. In some way we build phenomenal reality from empirical reality. The first level is then the phenomenal reality and the second one the empirical reality. Now, this second level is very peculiar. It is not composed of objects or strengths or fields or anything else that is representable. It is the set of all potentialities that can be actualised. This actualisation can be effective only if it respects some constraints preventing us to create the phenomenal reality as we want. Empirical reality is then the set of all these potentialities and the associated constraints. Now we can understand why contradictory theories can be empirically adequate (what we should now call phenomenally adequate). It is because these theories respect the constraints imposed by empirical reality.³⁴ The reason why they are empirically adequate is neither the fact that the terms of the theories have a real referent in empirical reality (which is not composed of physical entities) nor the fact that the referent is in the phenomenal reality which is only composed of actualised perceptions. There is no more reality left to which physical theories can refer. Physical theories seen as algorithms are adequate if their algorithms are in adequation with the structural constraints imposed by empirical reality.

8.2 The Need for a Third Level

Phenomenal reality is both conceptualizable and representable. Empirical reality is conceptualizable since we are able to build theories which reflect its structure but it is not representable. Its very nature (set of potentialities) is an obstacle to any representation since a representation is by definition actualised. Moreover, it is conceptualizable in many different ways because of underdetermination of theories. This means that it is only partially conceptualizable because it is impossible to gather all the different ways in a global conceptualisation. It remains beyond every exhaustive description. This is an analogue to quantum complementarity. In quantum physics, it is possible to measure the position of an electron or its momentum but not both simultaneously. Moreover measuring first position then momentum doesn't make it possible to know both values afterwards. Phenomenal reality is a sort of section of empirical reality, coming from many measures and giving a partial representation of it. Each section is exclusive in the sense that it is impossible to rebuild a global view of empirical reality through different sections as is the case for example in architecture where 3D pictures are drawn from different 2D views. Thus, that empirical reality is not representable doesn't come from the fact that some of its parts are out of reach but from the fact that it is impossible to have a total and global knowledge of all these parts simultaneously.

One can wonder whether these two levels exhaust “the universe”. It seems to me that it is naïve to think that “everything” is conceptualizable. Of course we must be very prudent here. If by “what is conceptualizable doesn’t exhaust what is in the universe” we mean “there is something that is not conceptualizable” we fall immediately inside a trap of language. Let’s remember Wittgenstein’s famous sentence “Whereof one cannot speak, thereof one must be silent”. I am here going to try to show what I mean more than to describe it precisely. Empirical reality is the set of all actualisable potentialities and is a sort of asymptotical construction built from all the concepts that we use to describe and to predict phenomenal reality. Then in a way, it is linked to the capacities of the human brain. It is obvious that this empirical reality is totally out of reach for a dog or for a monkey. I pretend that our capacities of conceptualisation are limited and I refuse to consider that for example, some greater capacities are impossible. That means that for entities having these superior capacities, some things that are conceptualizable for them are not for us, exactly in the same way that what is conceptualizable for us is not for a dog. Thinking the contrary would mean that the human brain is the ultimate machine for conceptualising which seems a little bit pretentious. Then I will postulate that there are “entities” that are not conceptualizable. Language finds here its own limits because it is not possible to say anything about these “entities” which are not usual things. Let’s only say that conceptualizable things are not everything and that there is something else which I’ll call the third level. Thus there are three levels: the first one is the phenomenal reality that is representable and conceptualizable, the second one is the empirical reality that is conceptualizable but not representable and the third one (for which I give no name) is not conceptualizable.

9 Conclusion

This stance is compatible with the results of quantum mechanics and their interpretation that undermines traditional realism. It rejects metaphysical realism and epistemic realism but is not an idealist position. It also enables us to understand why apparently contradictory theories can be empirically adequate without being trapped in the false questions of the existence of a real referent of theoretical terms or of the wrong concept of approximate truth. Underdetermination of theories becomes a natural consequence of the fact that theories are algorithms used to predict phenomenal reality. The “no miracle” argument vanishes in front of the fact that our theories work only inside the part of phenomenal reality that lends itself thereto. The success of these theories comes from the fact that they respect the structural constraints of empirical reality. Then the fact that “something says no” comes from the difficulty for our mental structures to work out a theoretical formal construction and a perceptual construction which are jointly consistent. Due to the empirical blindness features, it is certainly a sceptical epistemology but it doesn’t refuse to recognize the possibility of obtaining some pragmatic knowledge about phenomenal reality.

This stance borrows some features from realism in that it refuses to say that everything comes from our mind but at the same time it doesn't admit the existence of an independent reality. Phenomenal reality and empirical reality exist only relatively to our perceptive capacities. Their existence is then of a different nature than the one postulated in traditional realism. The Human mind plays an essential role but is not the sole ingredient. As Putnam says³⁵ "the mind and the world jointly build the mind and the world".

Endnotes

- 1 Agazzi, 1989, p. 91.
- 2 Ibid.
- 3 Van Fraassen, 1980.
- 4 Sellars, 1963.
- 5 Rescher, 1987, p. 2.
- 6 Actually, there are 5 different string theories which can be shown to be equivalent inside an eleven-dimensional M-theory.
- 7 Putnam, 1987, p. 3.
- 8 Agazzi, 1989.
- 9 D'Espagnat, 1994.
- 10 D'Espagnat, 2002.
- 11 Boyd, 1989.
- 12 When two quarks become separated, as occurs in collisions made in particle accelerators, a new quark/anti-quark pair emerges out of the vacuum. So instead of seeing the individual quarks, physicists see what they call "jets" of particles.
- 13 Though this is still a controversial debate, see the analysis in d'Espagnat, 1994, 2002, Zwirn 1997, 2000, to appear and Soler, 2006.
- 14 I have no place to develop it here (see Zwirn, 2000).
- 15 I must admit that this is not a rigorous proof but rather a reasoning by analogy.
- 16 D'Espagnat, 2002.
- 17 Gardenfors, 1988.
- 18 Boyd, 1989.
- 19 Rescher, 1987, p. 73.
- 20 In any case, in the absence of any plausible definition, this point is of little interest.
- 21 Putnam, 1982.
- 22 See in particular the discussion about the truth in realism in Newton-Smith, 1989.
- 23 This is not totally correct because the program has to contain also the instruction "write" and the string needed to delimitate it, but the longer the string, the smaller the difference.
- 24 Provided that suitable initial conditions are given.
- 25 Chaitin, 2005, p. 53.
- 26 Hanson, 1958.
- 27 Kuhn, 1962.
- 28 Feyerabend, 1975.
- 29 Van Fraassen, 1980.
- 30 We have previously refused to count the prediction of a novel fact as an argument for scientific realism but we consider that such a prediction is an argument for the empirical pertinence of a theory because it proves that the theory is not a sterile construction limited to a description of recorded data.
- 31 Boyd, 1989, p. 8.

- 32 For lack of place here, it is not possible to explain why it is necessary to use as complicated a theory as ZFE. For more details about it see Zwirn, 2000 where the detailed reasoning is drawn.
- 33 Dummett, 1978.
- 34 I have not place here to develop this point which is related to structural realism. See Zwirn, 2000.
- 35 Putnam, 1981.

Bibliography

- Agazzi Evandro (1989), "Naïve Realism and Naïve Anti Realism", *Dialectica*, vol. 43, pp. 83–98.
- Bitbol Michel, Sandra Laugier (1997) (eds.), *Physique et réalité (Un débat avec B. d'Espagnat)*, Paris, Éditions Frontières-Diderot.
- Boyd Robert Thomas (1989), "What Realism Implies and What it Does Not", *Dialectica*, vol 43, pp. 5–29.
- Chaitin Gregory (2005), *Meta Math! The Quest for Omega*, New York, Pantheon Books.
- Dummett Michael (1978), *Truth and others Enigmas*, Cambridge, Harvard University Press.
- Espagnat Bernard d' (1994), *Le réel voilé*, Paris, Fayard.
- Espagnat Bernard d' (1997), "Physique et réalité", in M. Bitbol et S. Laugier, 1997.
- Espagnat Bernard d' (2002), *Traité de physique et de philosophie*, Paris, Fayard.
- Feyerabend Paul (1975), *Against Method*, London, New Left Books.
- Gärdenfors Peter (1988), *Knowledge in Flux: Modeling the Dynamics of Epistemic States*, Cambridge, Massachusetts, MIT Press.
- Hanson Norwood Russel (1958), *Patterns of Discovery*, Cambridge University Press.
- Kuhn Thomas (1962), *The Structure of Scientific Revolutions*, University of Chicago Press.
- Newton-Smith William (1989), "The Truth in Realism", *Dialectica*, vol. 43, pp. 31–45.
- Putnam Hilary (1981), *Reason, Truth and History*, Cambridge University Press.
- Putnam Hilary (1982), "Three Kinds of Scientific Realism", *Philosophical Quarterly*, 32.4, pp. 195–200.
- Putnam Hilary (1987), *The Many Faces of Realism*, La Salle, Ill., Open Court.
- Rescher Nicholas (1987), *Scientific Realism*, Dordrecht, Reidel.
- Sellars Wilfrid (1963), *Science, Perception, and Reality*, Atlantic Highlands, NJ, Humanities Press.
- Van Fraassen Bas (1980), *The Scientific Image*, Oxford, Oxford University Press.
- Soler Lena (2006), *Philosophie de la Physique*, Paris, L'Harmattan.
- Zwirn Hervé (1997), "La décohérence est-elle la solution du problème de la mesure?", in M. Bitbol et S. Laugier, 1997, pp. 165–176.
- Zwirn Hervé (2000), *Les limites de la connaissance*, Paris, Odile Jacob.
- Zwirn Hervé (to appear), "Can we Consider Quantum Mechanics to be a Description of Reality ?", in *Rethinking Scientific Change and Theory Comparison. Stabilities, Ruptures, Incommensurabilities*, L. Soler, P. Hoyningen, H. Sankey (eds.), Vienne, Springer.

Philosophy of Chemistry

Bernadette Bensaude-Vincent

The notion of “philosophy of chemistry” challenges the singular in the phrase “philosophy of science”, which is the standard term for the discipline in the English language. This linguistic peculiarity has undoubtedly favored the tacit equation science = physics that has characterized mainstream philosophy of science during the course of the 20th century. The hegemony of physics has had profound consequences that have subsequently become identifiable. One of them is the increasing gap between philosophical reflection and science in action. As Joachim Schummer has pointed out: “Had those philosophers without prejudice gone into the laboratories, then they would have stumbled on chemistry almost everywhere”.¹ For there is a striking contrast between the philosophers’ neglect of chemistry and the quantitative data, which show that chemistry is by far the largest scientific discipline in terms of the number of publications indexed by the major journals of abstracts. Thus, philosophers have virtually ignored the major part of scientific activity choosing instead to focus on theoretical physics, which seemed more appropriate in light of the “linguistic turn”.

The situation is slightly different in the European tradition. The plural “*philosophie des sciences*” which has prevailed in the French language may be due to Auguste Comte’s longstanding influence, since he strongly advocated a regional epistemology. The result is that chemistry has not been totally neglected. As I have argued elsewhere, chemistry helped shape the French tradition, especially in what can be labeled its “historical turn” and its focus on theories of matter.² Whether French philosophers interacted more with active scientists than their Anglo-Saxon counterparts or shared the scientists’ interests remains a matter of debate for historians of the philosophy of science.

After decades of neglect of chemistry in mainstream philosophy of science, however, the late 20th century witnessed an impressive revival of philosophical interest

B. Bensaude-Vincent (✉)
Département de philosophie
Université Paris X
200 avenue de la République
92001 Nanterre, France
E-mail: bensaude@u-paris10.fr

in the discipline. Philosophy of chemistry has become a dynamic research field, establishing itself as a sub-discipline in the 1990s. An *International Society for the Philosophy of Chemistry*, founded in 1997, has organized an annual summer conference. Two journals have been launched, *Hyle* in 1995 and *Foundations of Chemistry* in 1999. Chemists and chemistry teachers have been the prime movers behind this renaissance of the philosophy of chemistry. For them, the hegemony of physics in the philosophy of science resonated with the reductionist ambitions of quantum physicists, who denied the very existence of any independent theoretical foundations of chemistry. For chemistry teachers, Paul Dirac's famous 1929 claim that "the underlying physical laws necessary for the mathematical theory of a large part of physics and the whole of chemistry are thus completely known", had always been a trauma, as it meant that their discipline could be taught as a sort of applied physics.³ Their concern with the philosophical implications of Dirac's statement was not shared by working chemists, who knew that the reductionist research agenda was impossible to achieve because the calculations would always be too complex. However, when digital computers allowed *ab initio* calculations, theoretical chemists started to worry once again about reductionism and became more interested in philosophy. Chemists felt the need to demonstrate that chemical concepts could not be deduced from quantum mechanical principles, giving rise to a flood of technical publications about reductionism in the 1980s.⁴ Chemists, advocates of the autonomy of their discipline, tend to use philosophy as a battlefield for their heroic struggle against the imperialism of quantum physics. As a consequence, reductionism and foundational issues have been the main concern over the last decade. Subtle conceptual distinctions became strategic to limit the dominion of quantum mechanics over chemistry: "quantitative reduction" does not mean "conceptual reduction", "ontological dependence" does not imply "epistemic dependence".⁵ The notion of supervenience referring to asymmetric dependence has been envisaged as a possible substitute for the notion of reduction.⁶ In this perspective, only a few aspects of chemistry – such as the interpretation of the periodic system – have drawn philosophical attention, while concepts and practices in daily use by laboratory chemists have been overlooked. Ironically the overwhelming concern with reductionism threatens to lead to a reduction of the emerging field of philosophy of chemistry to theoretical issues. If this trend continues, chemistry would paradoxically be bound in philosophical allegiance to physics, condemned to spend its existence ruminating over Dirac's arrogant claim.

It is time for philosophers to face up to what is the most evident feature of chemistry, that it is not only a natural science but also a cornucopia of material technologies. Explaining and modeling are just two of its many facets. Chemistry is also about making, testing, measuring, improving yields... The dual face of chemistry demands a specific philosophical approach. It is not enough to revisit philosophical notions that have been sanctified in the context of a tradition of philosophy of science that has modeled its categories around theoretical physics. Indeed, to try to accommodate these notions to chemistry understood in its entirety is a hopeless task.

Chemistry needs a philosophy of its own. A number of French philosophers—Pierre Duhem, Émile Meyerson, Hélène Metzger—have paved the way for such an

approach. In particular, Gaston Bachelard has suggested an alternative philosophy that he termed “metachemistry”.⁷ Bachelard’s aim in *The Philosophy of No*, was to describe a new trend in science embracing non-Euclidian geometries, non-Aristotelian logics, non-Cartesian physics, non-Lavoisieran chemistry. The prefix “non” means (i) that today’s science is not the continuation of the past and rather questions and challenges established knowledge through a polemical process; (ii) the non-sciences are not however negations of past theories and rather include them as particular cases in a dialectical process. Since Bachelard’s aim was to promote a new updated “scientific spirit” rather than digging into the singularity of the philosophy of chemistry, here I will try to explore what Alfred Nordmann presented as “the promise of metachemistry”.⁸ This aim cannot be achieved without a move away from the “linguistic turn” that has prevailed in the logical-positivist tradition, aligning instead with the “practical turn” that characterizes more recent philosophical trends.⁹ In this respect, I am merely following a path opened by Roald Hoffmann, Nobel laureate for chemistry in 1981, and adding a historical dimension to his philosophical essays.¹⁰ I will first consider the impact of laboratory practices on chemical explanations and theories, before turning my attention to the issue of the ontological burden of chemistry.

1 Knowing Through Making

Any philosophical examination of chemistry should take into account the fact that chemistry is and always has been a laboratory science. The word “laboratory” itself, originally referred to the place where chemists worked and only gradually spread to include the spaces used for other kinds of experimental practice. Frederic Larry Holmes, a leading historian of 18th century chemistry, insisted on the importance of this physical setting:

The problems and objects of study of chemistry have been provided by and limited by the operations that could be performed on materials in a chemical laboratory [...]. As theoretical structures changed and new objectives supplemented or displaced older ones, the stable setting of the chemical laboratory both identified chemists and distinguished them from other natural philosophers who dealt with some of the same phenomena that concerned them.¹¹

This physical niche determines both the object of chemical investigation and a specific way of knowing that is the chemists’. As the etymology of the term reminds us, the laboratory is a place of labour, of manual work rather than of inductive or deductive reasoning. The practice of chemistry is as much a physical activity as a mental exercise. Joan Baptista Van Helmont used to say that “God sells the arts in return for sweat”, meaning that knowledge of nature was to be obtained only at the cost of painstaking experiments.¹² Chemists attempt to know substances by transforming them by means of manipulations and physical operations. Whatever the importance of chemical theory, chemistry is first and foremost concerned with making. Historically it was an art and craft before it became an academic science.¹³

Nowadays, if we look at scientometric studies, we can see that making new molecules remains a major part of the work found in academic publications.¹⁴

Historically, chemistry provided the grounds for criticizing the *esprit de système*, embodied by scholars “speechifying” in their doctoral robes. As an illustration, we can cite Diderot’s blistering offensive launched against speculative and abstract knowledge in *De l’interprétation de la nature*, an attack echoed in Gabriel François Venel’s heroic portrait of the chemist as an “artist”, in his article “*chymie*” in the *Encyclopédie*.¹⁵ More recently, Roald Hoffmann has written:

The reliable knowledge gained of the molecular world came from the hot and cool work of our hands and mind combined. Sensory data, yes, but we did not wait for Scanning Tunneling Microscopes to show us molecules; we gleaned their presence, their stoichiometry, the connectivity of the atoms in them and eventually their metrics, shapes and dynamics by indirect experiments.¹⁶

Indeed, “indirect” may be a key word for understanding the chemists’ way of knowing nature. They use the detour of the laboratory to access nature. This does not simply mean that they use the mediation of instruments to understand natural phenomena, like experimental physicists do. Rather, they take mediating practices much further by insisting that only man-made, artificial products provide information about natural substances. To know the nature and properties of substances, chemists proceed by analysis and synthesis. Since the Renaissance, decomposition or the resolution of bodies into their components, combined with recombination or the recombination of the purported components to give the original substance, has provided the key to understanding material substances.¹⁷ Joan Baptista Vico’s famous statement *Verum et factum convertuntur*, established that we can get rational knowledge only about what we have done.¹⁸ For chemists, we can know only what we have produced through technological processes. As Bachelard noted, even when they extract plants or minerals from nature, chemists first submit them to a number of purifying processes.¹⁹ Thus, they rely on facticity to understand nature. This is how Bachelard interpreted Marcelin Berthelot’s famous statement: “Chemistry creates its object”.²⁰ Making things and making them as pure as artefacts is the chemist’s approach to nature.

Bachelard also emphasized the asymmetry between analysis and synthesis. Indeed analysis can provide chemists with some *evidence* about the nature and proportion of the constituents of substances. However, it will never give them *confidence*, for there remains the suspicion that the results of analysis were produced by the analytical tools rather than being preexistent in the compound. Analysis lacks definitive demonstrative power. While it may serve the purposes of falsification, only synthesis has the power to confirm. There is no way to overcome objections apart from recomposing the original compound from its purported components. Synthesis thus stands as the realization of a conjecture about the composition or the structure of a substance. Chemical proofs depend on the reciprocity of analysis and synthesis, which are both indissociably intellectual and experimental processes. Their reciprocity is at the root of Immanuel Kant’s admiration for Georg Ernst Stahl, who “transformed metals into calx and calx into metals”.²¹

2 “Making up Stories While Making Molecules”²²

Making is the chemists’ major activity, and it is more than simply a material practice. It also characterizes an intellectual practice. As Hoffmann put it, they are “making up stories” about what they are doing with their hands and flasks. Chemical theories, unlike theories in physics, are not really aimed at explaining phenomena. Rather, they try to make sense of phenomenological data using stories about tiny invisible atoms or molecules. As early as the 17th century, Nicolas Lemery forged hooked and spiny atoms to account for the behaviour of acids and alkalis, while modern chemists use molecular models to predict new compounds. In so doing, they do not claim to provide a causal explanation, and their theory is closer to being a narrative. Just as early-modern hooked and spiny atoms were a “Cartesian novel”, modern electronic orbitals could be regarded as a “quantum novel”. Similarly, the structural formulas invented by 19th century chemists were not meant as representations of the real world of atoms and molecules. Thus, Charles Gerhardt, who was a staunch advocate of atomic notation, drew the formulas of organic compounds according to three molecular “types”. He used these types to interpret a great many reactions, and even predicted unknown compounds by substituting radicals for hydrogen in each of the types. But he never suggested that his formulas reflected the internal architecture of the compounds he was representing and refused to view the radicals as isolable and real bodies. They were useful and indispensable fictions.

Nevertheless, speaking of “fictions” does not necessarily mean that chemical theories have no truth-value at all or that they should be viewed as mere instruments for prediction and classification.²³ Instead we need to redefine what counts as the truth-value of chemical statements. The dilemma of instrumentalism (or positivism) versus realism is a pitfall that chemists need strenuously to avoid. If by realism we mean the representation of an external reality, it is just as inadequate a label as instrumentalism. Chemists make extensive use of visual images but these are not intended to refer to real individual molecules. Rather they are better thought of as icons representing relations between individual entities. Chemists seem to share the conviction that the bedrock of chemical properties does not lie at the ultimate level of matter. In other words, they do not strive to reach the roots, or to unveil the ultimate building blocks of matter. They make up plausible narratives to account for the properties observed in individual substances that they use, or to predict and make new substances with desired properties. In so doing, they are constantly shifting from the macro- to the micro-level.²⁴ Thus, they never settle on a scale for their reflection, with the constant shifting between levels determining their characteristic expository style. Chemistry textbooks, whether from the 17th century or most recent ones, tend to juxtapose narratives of experiments performed at the macro-level with narratives about relationships between microscopic invisible entities. The two kinds of narrative run in parallel but neither alone accounts for the ultimate causation.

Rather than being ideal accurate representations of nature, these narratives display meanings, with atoms and molecules best described as actors in a story. Even

when these invisible entities are visualized using imaging techniques, they do not mirror the ultimate reality underlying phenomenological appearances, although they do *mean* something for the chemists. In certain cases they may mean that there is a possibility of breaking a bond, or of substituting a functional group or of encapsulating certain atoms within a cage molecule, etc. In addition stories require a temporal structure: temporality plays a prominent role in chemical narratives as the kinetics determines whether the reaction will be a success story or not. Wilhelm Ostwald was, like Berthollet, concerned with incomplete reactions whose outcome depends on subtle equilibriums, and proposed new narratives of chemical experiments based on the frequency of collisions. Thus, for example, catalytic materials that prompt the advancement of a reaction in a specific direction play a similar role to that of the hero's companion in epic narratives.

3 Requirements and Obligations

Hoffmann's metaphor of story telling suggests that chemical theories have very weak explanatory power. In fact, Hoffmann makes the case for the "power of poor theories" and insists that two alternative theories, belonging to different paradigms, are not necessarily incommensurable.²⁵ A standard example is Linus Pauling's theory of the chemical bond, associating Lewis's notion of a shared electron pair with the quantum mechanical notion of a covalent wave function, which proved to be extremely useful in heuristic terms. Hoffman comments:

I think incommensurability is no problem whatsoever to chemists. Differences in language are there, the result of different paradigms, but more so of history, and of education. Yet people, eager to make things, with no handwringing on how problematic it all is, graft one way of understanding onto another.

Making up stories does not, however, mean that chemists rely on fanciful and arbitrary accounts. It just means that chemists do not claim to reach the roots, or the ultimate cause of phenomenological data. Chemistry, like other experimental sciences, is a normative activity. But if its ruling norm is not to provide *the* perfect representation of reality, we may nevertheless demand what kind of norms are in use in this science.

The distinction between requirements and obligations forged by Isabelle Stengers in her "ecological" approach to practices in science is particularly helpful for characterizing chemical practices.²⁶ Experimental scientists like to see their activities as conforming to a number of criteria or standards, including logical rules, experimental controls, peer review, etc. Conforming to such general widely accepted rules allows them to draw a clear demarcation line between their practices and others that are generally considered to be non-scientific or at least less scientific. Fulfilling such criteria is thus indispensable for defining the identity of a scientific practice, and in this respect chemists are no exception. They comply with the canons of the so-called scientific method, which shows that they are full members of the scientific tribe. However, Stengers argues, experimental practices are also governed by

a number of more elusive and tacit norms – dubbed “obligations” – instituted by active scientists in specific contexts. The chemists’ obligations are the collective standards that they have adopted over the centuries in order to learn something about nature from their experimental practices, while at the same time never forgetting about making, and producing new artefacts or drugs.

Galileo’s major obligation – that only matters-of-fact can tell us about the truth – led him to question nature in a mathematical language in his experiments on falling bodies. By contrast, the chemists’ major obligations seem to be caution and skepticism. Crystals, liquids or gases in flasks behave in unpredictable and sometimes positively dangerous ways. These behaviors are so puzzling that chemists had to forge an arsenal of obligations: purifying, synthesizing, submitting them to standard reactants to settle their identity and characterize their properties. For this purpose of identification they rely on a wide range of different tests. In medieval times, chemists tested everything with fire, but since then they have come to use all sorts of chemical reactants and physical techniques of measurement, ranging from traditional balances, hydrometers, gasometers, to modern spectrometers. They have to pay particular attention to the conditions of their experiments – in some cases even more than experimental physicists – as slight modifications in temperature, pressure, concentration etc., can alter the course of a reaction, thereby changing the composition of the product. Since Robert Boyle’s famous publication, “sceptical” has often been associated with the word “chemist”. It does not mean that chemists are stubborn unbelievers. Rather it is because what they know about chemical substances and chemical reactions justify a cautious attitude concerning any conclusions they might be tempted to draw from their experiments.

Identifying, naming, and classifying are the chemists’ principal responses to their major obligation. Due to their “creativity” – millions of new molecules are reported in the *Chemical Abstracts* each year – chemists are continuously under pressure, as they have to find a name and a place for all these newcomers in their databases. In 1787, when a group of French academicians designed a “method for naming”, they assumed that by formulating the major requirements for a chemical nomenclature, they would provide subsequent generations with reliable guidelines for naming any newly discovered substances.²⁷ They formulated general rules for coining systematic names based on composition, and banished names based on the substance’s qualities, its uses, or the circumstances of its discovery. In doing so, they were acting as “architects of matter”, designing and planning future chemical edifices. The growing number of organic compounds in the 19th century nevertheless generated a chaotic situation with dozens of different names for the same substance. Standardization and systematization were the two leading requirements reiterated at the end of the 19th century by the first International Conference on chemical nomenclature held in Geneva in 1892. The concerted response was to give each substance an official name, but most of them were never used by chemists in their daily chemical practices. Indeed, this ideal of standard and systematic names has been continuously challenged, and linguistic customs established within scientific journals tended to prevail, meaning that the standard names in common use no longer complied with the original ideal of a systematic nomenclature. Regular

international meetings and a permanent commission on nomenclature at the International Union for Pure and Applied Chemistry continue periodically to revise the rules. Still, the current nomenclature is by no means as systematic as the 1787 reformers had envisioned. Trivial names – names that do not refer to the structure of the compound – coexist with the systematic names that conform to the rules. In fact, both in organic and inorganic chemistry, most names are semi-trivial, mixing informal parts with those constructed following the systematic rules. Thus, the difficulty of keeping up with systematic names for extremely complex compounds proved so difficult that chemists had to renounce their ambition of submitting the molecular world to their ideal of rational systematization. This obligation may be considered a fundamental weakness, a sign of the imperfection of the chemical sciences. But what Stengers means by “obligation” suggests a more charitable reading. This term suggests a kind of binding agreement between chemists and the object of their investigations. Chemists are “obliged”, in the dual sense of the word, bound by and indebted to the growing population of molecules they both create and investigate. They are less “architects of matter” than dusty laborers trying to discipline a jungle of diverse molecules.

The repeated attempts to classify chemical elements during the 19th century provide another illustration of the interplay between requirements and obligations. The official requirement was to group simple substances according to their common properties. Nevertheless, chemists soon realized that the ideal of a “natural classification” reflecting all the similarities between the elements would be impossible to achieve. They consequently adopted “artificial classifications”, based on one or two properties arbitrarily selected among the wide variety of candidates. They even combined artificial classifications for metals with natural classifications for non-metals while admitting that the division between metals and non-metals was itself artificial. Such hybrid arrangements are far from the rational ideal and might therefore be considered a major defect. Chemists were, however, “obliged” to adopt and teach such imperfect solutions, as they were aware that their picture of the material world was inevitably biased, that between the exigencies of an operational system and an ideal one something had to give. Emile Meyerson, a chemist turned philosopher, argued that although the distinction between metals and non-metals was arbitrary, chemists used it because they had to draw strong distinctions, to artificially introduce rigid demarcations into the flux of complex inter-relations, in order to be able to refute conjectures. Rigidity and falsification add truth-value to the story invented by chemists. Meyerson used a suggestive metaphor, borrowed from Arthur Balfour, to characterize the chemist’s approach to nature: they are “drawing a fiber” out of the magma of reality. Chemical classifications seem to be based on the assumption that nature is composed of a “fibrous structure” in which they select a specific region in order to disentangle the local network of relations.²⁸ Focusing on a fiber, they start reasoning about its connections with the whole fabric, while all the time looking at the landscape created by the extraction of this one fiber. They never claim that this fiber is the root of the structure, or the unique entry into the puzzle. But drawing out a fiber is their obligation, which means, on the one hand, that they must not break it, and that they must use it as a robust guideline. On the other hand,

they are not to treat the fiber as a completely secure element that would permit safe deductions. “If...then” is a forbidden leap in a jungle where unexpected surprises are strewn on every pathway. Thus, chemical classification remains an open field. More than a century after Mendeleev’s periodic system came to be considered “the chart of nature” a view subsequently justified by atomic physics, chemists are still unsure about the best way to represent the periodic function. There is no ideal chart. Each year, new systems are designed and new graphic representations are submitted for publication, some of them concocted by obscure chemical practitioners, suggesting that classifying elements remains a work in progress, a communal and endless task.²⁹

4 “No Natural Body Consists of Matter *Per Se*”

An aura of materialism surrounds the image of chemistry, which derives as much from the chemist’s concern with material things as from the abundance of material goods generated by the chemical industry.

Ironically, however, chemists do not care very much for matter. They have used the terms “substances” or “bodies” for centuries but, as Venel noted in Diderot’s *Encyclopédie*, “no natural body consists of matter *per se*”.³⁰ Rather than being concerned with matter in general, chemists want to know why only one particular acid dissolves gold or why spirit of niter joined to salt of tartar produces true saltpeter. They pay attention to individual properties, with reference to a jungle of different materials and their potentialities.

Chemistry is concerned with the stuff things are made of, but we need to ask what concept of substance they use. In *The Philosophy of No*, Bachelard argued that the metaphysical notion of substance inherited from the Ancient Greek quest for permanence has been modeled on classical physics. Since Descartes, matter has been regarded as an essentially homogeneous substance defined in geometrical terms, with the diverse sensory properties that characterize the multiplicity of the phenomenal world being merely “secondary qualities” arising from the spatial arrangements and rearrangements of indistinguishable elements. This metaphysical notion of substance as a permanent and pervasive substrate underlying phenomenological change is, however, completely inappropriate for chemistry. What might a “metachemical” notion of substance look like? The stuff that chemists call “substances” is always in the plural.³¹ For chemists, substances are concrete entities with individual properties. Explanations of chemical phenomena rely on a few immutable elements responsible for the individual properties of compounds. They may be irremediably invisible but they can be traced by means of the sensible effects that they cause at the phenomenological level, or by means of their circulation from one combination to another.

The dichotomy concerning this issue outlined above suggests that as far as the philosophy of matter is concerned, physics and chemistry are heirs to two different ancient traditions, with physics deriving from Democritus and Epicurus and chemistry from Empedocles and Aristotle. In the former case, the endless variety of

substances with their individual specific properties is referred to essentially similar atoms, only distinguishable with respect to their figures and movements, while in the latter case, the variety of individual properties is attributed to strongly individualized principles. In *Le Mixte et la combinaison chimique*, Pierre Duhem suggested a similar distinction between two “research schools”. The first was the “corpuscular school” – Cartesian and then Newtonian – which one could characterize by Boyle’s assumption of a “catholic matter”, and which would lead to the mechanistic models of the 19th century that Duhem rejected. The second was the “Aristotelian school”, taken to be characterized by its rejection of all “systems”, all *a priori* reasoning, as well as its firm attachment to irreducible qualities.³²

This dual genealogy is, however, superficial and in the end misleading. The so-called rival paradigms – the monist, atomistic, mechanistic philosophy versus the pluralist, qualitative doctrine of elementary principles – were not incommensurable. Most chemical theories managed to combine them in some fashion. As historians of early modern chemistry have shown, a corpuscular theory was embedded in the alchemical tradition, and was, in fact, crucial for justifying the possibility of transmutation.³³ It is now well established that Boyle’s corpuscular philosophy, for example, stemmed from this longstanding alchemical tradition transmitted via Daniel Sennert. Thus, Boyle’s corpuscular philosophy was not the grafting of a physical theory onto a previously incoherent body of alchemy or iatrochemistry.³⁴ Later on, Georg Ernst Stahl also assumed that material bodies were constituted by the *mixta, composita* and *supercomposita* of constituent particles. He assumed a corpuscular view of matter meshed with a view of individual principles acting as the vehicles of the properties. Such combinations suggest that atomist views and the principle theories were deployed for different purposes and did not address the same issues. Neither of them holds the secret of matter. For chemists, there is no privileged ultimate level of reality; instead they adopt what Bachelard termed a “laminated reality” since laboratory practice gives access to substances at multiple levels simultaneously.³⁵

5 The Chemists’ “Essential Tension”

Stahl used a clear-cut distinction to differentiate the territory of chemistry from that of physics. He acknowledged that mechanical physics could account for one species of material compounds, namely “aggregates”, whereas only chemistry could deal with “mixts”.³⁶ Aggregation was a juxtaposition of units, and could be understood in mechanical terms such as mass and movement. Mixtion, however, was the union of principles involving individual affinities. The decomposition of an aggregate would not affect the properties of its components whereas the dissociation of a mixt entailed changing the properties of its elements.

This conceptual distinction echoed the issue raised by Aristotle in *De generatione et corruptione I* about the mode of presence of the constituents in a mixt. The problem emerged from a critical review of atomism. If atomist doctrines were right, then a mixt would be just a collection of atoms placed side by side, like grains

of wheat and grains of barley. “To the eye of the Lynx nothing would be combined”³⁷ Constituents would be physically present in the compound although not visible at first glance. Thus, they can be recovered without changing the properties of the compound. Aristotle insisted that if the components are preserved unchanged then the mixt is only apparent. By contrast, a true process of mixture involves the interaction of qualitatively differentiated ingredients in such a manner that they do not persist unchanged in the resulting compound. A true mixt is not, therefore, composed of constituents sticking together. Something new is created, with properties not possessed by the original ingredients. The emergence of a new “stuff” implies that the ingredients no longer coexist with the mixt. Consequently, a true mixt can be characterized by an *either...or* condition. Either you get a compound and you lose the properties of the initial ingredients, or you recover the original ingredients and you lose the properties of the mixt. By contrast, the atomic conception of chemical combination does not demand such a disjunction.

Paul Needham, who offered a detailed analysis of Aristotle’s conception of mixts, has convincingly argued that Aristotle raised the fundamental issue of chemistry, i.e. the generation of new substances out of initial ingredients.³⁸ This clear recognition of the problem should not, however, be used to suggest that Aristotle conceived a “theory of chemical reaction and chemical substances”, as Paul Needham seems to argue, since chemistry did not exist as an identifiable branch of knowledge at this time.

Avoiding such anachronisms is important for grasping the concept that I have dubbed the chemists’ “essential tension”. By referring to the title of Thomas Kuhn’s famous book, I want to draw attention to the specificity of chemistry. Indeed the tension that Kuhn found implicit in scientific research between tradition and revolution, between conformism and iconoclasm is also at work in chemistry, although its identity has been shaped by a more specific tension between two competing views of chemical combination.³⁹ At the turn of the 18th century, Stahl’s distinction between aggregates and mixts was aimed at circumscribing a territory for chemistry, centered on the notion of the mixt, in a defense against attempts at annexation by mechanism. So successful was this conceptual strategy, that Stahl was proclaimed the founder of chemistry throughout the 18th century. A century later, however, chemists no longer used the word mixt, as the notion of composition prevailed. In particular, Lavoisier’s famous definition of elements as undecomposable substances was an integral part of a reorganization of chemistry along the lines of another distinction, that between simple and compound. Lavoisier, who came to earn the title of the founder of “modern chemistry”, redefined it as the science that aimed at decomposing natural bodies and “examining separately the various substances entering into their combination”.⁴⁰ To be sure the compositional perspective was nothing new, but with the reform of chemical language it became the dominant paradigm.⁴¹ In the new language, names of compounds were coined by simple juxtaposition of the names of their components, and were considered as “mirror images” of the actual composition of the material bodies in question.⁴² Lavoisier, who admired and extensively quoted Etienne Bonnot de Condillac’s *Logic*, adopted his views of languages as analytical methods as well as his notion of analysis as a

two-way process, from simple to compound and from compound to simple. According to Condillac, analysis is a mental process involving the successive visualization of the individual elements of a picture presented simultaneously as a whole to the senses. Condillac used the metaphor of sight-seeing from the window of a castle. Immediately I see a landscape, then by analysis the mind will distinguish and name each element of the landscape pre-existing in the global view.⁴³ Condillac's logic, inspired by algebra, in turn inspired Lavoisier's use of equations to describe chemical reactions. A compound is described as the addition of two constituent elements. It is entirely characterized by the nature and proportion of its constituents. The use of the sign "equals" in the equation clearly indicates that chemists are no longer thinking in terms of the either/or condition. The puzzling issue raised by Aristotle about the mode of the presence of ingredients in the compound has been laid aside, discarded rather than being solved.

Reinforced by John Dalton's atomic hypothesis, the compositional paradigm has proved very successful. By the middle of the 19th century, the definition of a compound according to the nature and proportion of its constituents was being challenged by a structural paradigm that emphasized the importance of the arrangement of the atoms in molecules. Nevertheless, empirical and structural formulas both eliminate the either/or condition. The actual presence of the constituent elements suffices to account for the properties of the compound.

Pierre Duhem's return to Aristotle's notion in the title of *Le mixte et la combinaison chimique* (1902) was clearly intended to undermine the prevailing atomist interpretations. The familiar example of sugared water in his introductory chapter summarized Aristotle's theory in a few words, and restored the legitimacy of the either/or condition:

What in general, then, is a mixt? Some bodies, the ones different from the others, are brought into contact. Gradually they disappear, they cease to exist, and in their place a new body is formed, distinguished by its properties from each of the elements that produced it by their disappearance. In this mixt, the elements no longer have any actual existence. They exist there only potentially, because upon destruction the mixt can regenerate them.⁴⁴

Duhem mainly reproached atomistic explanations for assuming that the properties of a compound could be deduced from those of its constituent elements or atoms. His criticism also encompassed Lavoisier's compositional paradigm, since elements are not conserved as such in chemical reactions.

Emile Meyerson indirectly addressed the same issue although, unlike Duhem, he claimed that chemists could not do without atoms. He nevertheless pointed to the either/or issue involved in chemical equations, starting with the observation that when chemists write the equation $\text{Na} + \text{Cl} = \text{NaCl}$, they obviously presuppose the conservation of matter.⁴⁵ He observed that, interpreted literally, a chemical equation is a non-sense. In asserting that the addition of a soft metal like sodium to a greenish gas like chlorine equates to a colorless salt, chemists seem to be oblivious of the very conditions of their laboratory practice. Although they continuously play on the potentialities of various individual substances and take advantage of their differences, they admit that the compound "equals" the sum of its initial ingredients.

Thus chemistry seems to be moved by two antagonist forces. On the one hand, chemists aim at reducing the qualitative diversity of substances to identity. They would like to deduce their empirical data from an ultimate hidden cause in order to “satisfy their rational tendency to identification”, to use Meyerson’s terminology. Chemical equations balancing the inputs and outputs of chemical reactions are the best expression of this effort aimed at identification. They presuppose subsistence throughout chemical change, or the conservation of elements in chemical reactions, even though the diversity of substances and their idiosyncratic behaviours constitute the very *raison d’être* of chemical practices. Without a diversity of substances with their own individual properties and without a diversity of processes of reaction, there would be no chemical reactions and so no chemistry. Thus, chemists have no choice but to face “irrationals” (again using Meyerson’s terminology). They sense that it is useless to try and reach the ultimate reality, and hopeless to try reducing everything to sameness.⁴⁶

The tension between the two conflicting views of chemical combinations is not necessarily to be understood as a fight between the rational and the irrational, or as a contrast between a rational tendency and a more pragmatic one. After all, atomic theories do not hold a monopoly over rationality.⁴⁷ Moreover, atomic notions, and molecular models are man-made “artifacts”, tools forged for theoretical and practical purposes. Nevertheless, the tension is an essential one, as neither of the perspectives is sufficient to account for chemical combinations, while the two descriptions do not work harmoniously together. Chemical combinations thus offer a new case of complementarity in Niels Bohr’s sense; two necessary but nevertheless exclusive descriptions of a phenomenon.⁴⁸

6 Matters of Concern

Because chemists are not really concerned with understanding the fine structure of matter, they have regularly dismissed all hypotheses concerning the real existence of atoms. For instance, August von Kekulé, who conjectured the hexagonal structure of benzene that formed the basis of most artificial organic compounds manufactured in the second half of the 19th century, denied the existence of atoms. More precisely, he banished the ontological issue from chemistry, claiming that it belonged to metaphysics. Thus, chemists made extensive use of atoms and molecular models while denying their existence or claiming that they were simply fictions. This apparently inconsistent attitude survived (in France at least) long after the first demonstrations of molecular reality and the founding of atomic physics. For instance, the French chemist Georges Urbain wrote in 1921: “It is not absurd to suppose that the atomic model is identical with absolute reality. However, we know nothing positively about it. This model is a work of art”.⁴⁹ Such claims have sometimes been viewed as evidence for the theory that, under the pernicious influence of Auguste Comte, French chemists were sticking to strictly positivist positions, and consequently lagging behind modern chemistry.⁵⁰ This apparently inconsis-

ent attitude was not, however, confined to the small circle of French chemists. The ontological status of bonds and orbitals was discussed at length by the founders of quantum chemistry, with some of them denying their physical reality in an effort to demarcate the chemical approach to concepts such as resonance borrowed from physicists.⁵¹ Chemistry thus appears as a science bound to ontological non-commitment, an attitude not shared by modern physicists.

If we resist the temptation of identifying the philosophy of physics as the “right model” for all of the sciences, how are we to understand the strange attitude of these non-committal chemists? For Meyerson, the chemists who denied the existence of atoms simply lacked authenticity.⁵² He assumed that all chemists professed a naive realism, a belief in the existence of things such as barium sulphide, for instance. Meyerson is right: chemistry is certainly not ontology-free, although he misunderstood its ontology. The assumptions underlying chemical practices do not concern things such as barium sulphide, or rather, to be more precise, this sort of “thingism” (*chosisme*) is not typical of chemists. Two major matters of concern more adequately characterize their ontology: (i) a concern for relations, and (ii) a concern for agency.

6.1 Relations

There is no question that chemists deal with individual substances and pay attention to their molecular structures, but these things are of interest to them only in so far as they enter into relation with other units. Nineteenth-century structural formulas were not meant to be images of reality, and yet nor were they pure conventions. Rather they depicted capacities for bonding, the so-called atomicity or valence. Similarly, series of compounds were essentially viewed as potential combinations or syntheses. Ernst Cassirer has emphasized the functional determination of the concept of atoms in *Substance and Function* where he convincingly argued that the treatment of an atom as the “absolute substrate” of properties is only apparent. In fact, the concept of atom serves as a mediator for mapping out a network of interdependent relations between objects.⁵³

Bachelard also emphasized chemists’ concern with relations rather than with substrates. Since relations imply at least two terms, chemistry necessarily presupposes various kinds of beings. The two features that Bachelard selected to define the rationalism of modern chemistry, which he dubbed “non-lavoisieran”, were that it was plural and relational. For him, Mendeleev’s system epitomized the shift from realism to rationalism, because “law prevailed over matter of fact”.⁵⁴

The focus on relations allows chemists to choose the unit of matter that best suits their views. For instance, in Pauling’s valence bond theory, atoms are the combining units, and their interaction results in the formation of molecules. By contrast, in Mulliken’s molecular orbital approach, the atom is no longer the relevant concept for understanding chemical bonds. Molecules are taken as the basic building blocks, formed by feeding electrons into molecular orbitals.⁵⁵

After quantum chemistry had drawn physics and chemistry into cooperation, chemists continued to debate about the ontological status of relations themselves. In this context, we can cite the debate that took place between G. W. Wheland and Pauling about double bonds and resonance.⁵⁶ Thus, time and again, chemists set themselves apart by rejecting the physical meaning of the concepts they are using. They champion artificiality or “facticity” not only in their experimental practice but also in their intellectual practice. If today’s chemists are no longer non-committal, it is mainly because they assemble atoms and molecules like Lego blocks. They believe in orbitals as long as they can explain things with them and design reactions. According to Hoffmann “The reifying power of synthesis, when you do it with your hands, time and again, is incredibly strong”.⁵⁷

6.2 Agencies

The chemists’ “things” are implicitly described in terms of structures, properties and functions. Molecular structures are above all conditions for the emergence of properties, which themselves are viewed as dispositions for desired performances. While chemists do not care for matter, they are by contrast always searching for materials, i.e. substances useful for something. Thus, in the 18th century, Hermann Boerhaave and Guillaume-François Rouelle redefined the four elements in terms of agents, conceiving them as both the constituent units of compounds, responsible for the conservation and transport of individual properties through chemical change, and instruments of chemical reactions. Rouelle introduced his four-element theory under the heading “Instruments” that included “natural instruments” – fire, air, water and earth –, and two artificial instruments – menstrua and vessels. The ancient radical distinction between nature and human artifacts was thereby being blurred in favor of an instrumental view of matter as an active operational process. Material principles were always at work, circulating from mixt to mixt, whether in laboratory vessels or in the depths of the earth or the heights of the heavens. Subsequently, following the rise of the compositional paradigm after the reform of chemical language, and later the structural paradigm linked with the emergence of organic chemistry, chemical names and formulas have been mainly used as “paper tools” for predicting operations and substitutions.⁵⁸ They display the possible uses of compounds through their structure. This action-oriented language inspired Bachelard’s description of structural formulas as “rational substitutes”, providing a clear account of the possibilities for experimentation.⁵⁹ This is why 19th century chemists could reject all ontological commitment concerning atoms and molecules, while using them like plumbers use pipes, valves, and joints. Even today chemists refuse to endow the atomic theory with the power of representing the world, as long as they are concerned with powers for intervening in the world. Atoms and molecules are just potential actors in the drama of chemical transformation.

Ian Hacking’s reflections on the way the physicists use electrons in electron microscopy is similar to the way chemists view the constituents of matter.⁶⁰

Electrons are less explanatory notions than instruments for acting or creating phenomena. Hacking's distinction between "realism about theories" and "realism about entities" can thus be applied to chemistry. To be sure, chemists are realists; they believe in the reality of entities, which allow them to operate on the outside world or to be affected by it. "Operational realism" would thus be the right phrase to characterize the chemists' philosophy. The material world is a theater for operations; the entities underlying observable macroscopic phenomena are above all agents.

In this respect, the three categories of structures, properties, and functions are not the most appropriate for the philosophy of chemistry. Aristotle provides better resources by the addition of his notion of potential, which remains appropriate for characterizing the modality of constituent elements in combinations.⁶¹ The dual nature of chemistry – science and technology – requires the whole panoply of subtler distinctions found in Aristotle's treatise on *Categories*, 8.⁶² Properties belong to the category of quality, but there are many varieties of qualities. States (for instance, hard or soft) differ from dispositions. The former are stable and durable "possessed" qualities, whereas the latter are ephemeral and easily altered. Both possessions and dispositions being acquired in specific circumstances differ from natural capacities embedded in the subject. They all differ from "affectations" (bitterness, sweetness), which simply refer to sensory properties. The chemists' art of synthesis takes advantage of the whole spectrum of capacities in order to put molecules to work, to make the molecules do what chemists cannot do with their own hands.

In 2003, Susan Linquist, a biologist from MIT's Whitehead Institute, announced at a conference that: "about 10,000 years ago, [humans] began to domesticate plant and animals. Now it's time to domesticate molecules".⁶³ But domesticating molecules is what chemists have been doing for centuries. At the cost of repeated experimental trial and error, they have managed to tame an incredible number of molecules, to get sufficient control over their reactions to be able to use them as agents for performing specific tasks. Nevertheless, this domesticated stuff has never worked in the same way as man-made tools or machines. Substances operate according to their own nature, even when they are chemical "creatures". Through a number of more or less spectacular chance events and deplorable accidents, chemists have learned that they are still at the mercy of unexpected outcomes and that reactants do not always behave in a foreseeable way.

In addition, chemists usually work with huge populations of molecules in their vessels. Unlike nanoscientists, who are trying to domesticate molecules one at a time, chemists have no control over individual molecules, although they may know a good deal about the species of molecules in question, especially when they have created the substance themselves. Nevertheless, the shift in scale of the operations has radically changed the relationship between men and materials. The slogan of the nano-initiative, "shaping the world atom by atom" expresses the ideal of control and full command that lies behind nanotechnology. Individual molecules are supposed to be reliable entities, responding predictably to precise

signals. So deep is the contrast between this culture of precision and the more crude tradition of chemists, that for Eric K. Drexler, a champion of nanotechnology, chemical synthesis is an inexplicable enterprise, which he compares to trying to assemble a car by putting all the necessary parts in a large box and shaking them up together.⁶⁴ Nevertheless, such miraculous processes constitute the everyday functioning of the world's chemical factories. The "cars" that the chemists have managed to assemble by such ham-fisted methods are new things, with the constituent parts no longer accessible or even visible. When deploying their art, that of making molecules work for them, chemists are not like Plato's demiurge, who builds up a world by imposing his own rules and rationality on passive matter. Rather, they are like a ship's pilot at sea, who, constrained by the force of the ocean and atmosphere, is obliged to channel or guide the forces and processes given by nature, and ultimately exhibits the powers inherent in nature in the outcome.

In guise of a conclusion, I want to offer a few reflexive remarks on the functions of history in this philosophical essay on chemistry. In his paper on "The relations between the history and the philosophy of science", Kuhn argued that bringing them together could be subversive, because philosophy and history were two distinct mental sets like the rabbit and duck in the famous Gestalt "duck-rabbit" figure.⁶⁵ Although this mutual exclusion seems quite alien to French scholars trained in a tradition that promotes the conviction that: "there is no epistemology that is not historical"⁶⁶ the functions of history in this essay have to be clarified. History is not a source of examples that serve to illustrate and confirm philosophical claims about the "essence of chemistry". There is no such thing as an immutable essence of chemistry that would fit this kind of strong philosophical program anyway. Instead, history is used here as a source of problems. The historical materials are not meant to allow us to reconstruct the past, rather they are an indispensable detour for grasping the problems at stake and the philosophical views shaped by chemists themselves in their investigative and productive practices.⁶⁷ For chemistry *is* a historical process. The journey into chemistry proposed in this essay should be thought of like a trip on a rocket ship that is continuously in motion, but changes direction in response to its environment and other circumstances, although overall retaining a more or less direct trajectory. The purpose was to identify the kind of problems and projects that have guided generations of chemists in defining this trajectory over time, thereby (unconsciously) reconfiguring the identity of their science.

Centuries of chemical practices oriented towards cognition and action have generated a set of specific obligations, which can be characterized as both epistemological and ethical rules. Caution, utility, and efficiency have been as highly valued as the quest for truth in the sense of *adaequatio rei et intellectus*. The chemical sciences are not aimed at unveiling the underlying reality beneath the surface. Instead, they deal with a jungle of molecules and strive to take advantage of their dispositions. Chemists are put under an obligation by these substances, by their structures, properties and capacities, meaning that respect, as much as responsibility, should be at the base of a chemist's ethics.

Acknowledgements I am very much indebted to Roald Hoffman, Isabelle Stengers and Jonathan Simon for their comments or revisions of earlier versions of this essay.

Endnotes

- 1 Schummer, 2006, p. 21.
- 2 Bensaude-Vincent, 2005.
- 3 Simões, 2002.
- 4 See for instance Lévy, 1979, Primas, 1983, Liegener & Del Re, 1987, Scerri, 1991, 1994, 1997.
- 5 Primas, 1983, Scerri, 2000.
- 6 Scerri, 2000.
- 7 Bachelard, 1940.
- 8 Nordmann, 2006.
- 9 Schatzki et al. 2001.
- 10 Hoffmann, 1995, Grosholz, Hoffmann, 2000; see also Berson, 2003.
- 11 Holmes, 1993, p. 478.
- 12 Van Helmont, J.B. *De febribus, Opuscula* chap. 15, 26, p. 58, quoted in Newman & Principe, 2002, p. 180.
- 13 Bensaude and Stengers, 1993.
- 14 Joachim Schummer's survey of 300 papers on synthesis from 1980 to 1995 concluded that most of them were aimed more at syntheses rather than at classification or theoretical reflection (Schummer, 1997).
- 15 Diderot, *De l'interprétation de la nature*, 1753. Venel, Gabriel-François article "chymie", in Diderot, d'Alembert eds. *Encyclopédie ou Dictionnaire raisonné des sciences, des arts et des métiers*, vol. 3, 1753, p. 420b. Diderot, *De l'interprétation de la nature*, 1753.
- 16 Hoffmann, 2005.
- 17 In the Renaissance, chemistry was often referred to as spagyria or spagyric art, a term which, according to a dubious etymology spread in 17th century chemical textbooks, derived from the Greek *span* (to pull apart) and *ageirein* (to put together), Newman and Principe, 2002, p. 90.
- 18 Literally "The true and the made are interconvertible", Vico, J.B., 1725.
- 19 Bachelard, 1952, p. 22.
- 20 Berthelot, Marcellin, 1867, p. 275.
- 21 Kant, *Critique de la raison pure*, Preface to the second edition, transl. Tremesaygues and Pacaud, PUF, 1968, p. 17.
- 22 Hoffmann, 2005.
- 23 Duhem, however, viewed chemical theory as a kind of natural classification.
- 24 Grosholz and Hoffmann, 2000, analyzed this shift from macro and micro in terms of a motion between symbolic and iconic representations.
- 25 Hoffmann, 2005.
- 26 Stengers, 2006, pp. 67–70. I am very indebted to Isabelle Stengers for this section.
- 27 Crosland, 1962, Bensaude-Vincent, 2003.
- 28 As Meyerson put it in a letter to André Lalande dated 1930: "Ce terme de fibre montre d'ailleurs clairement de quoi il s'agit en l'espèce; il faut déterminer, dans la masse confuse et troublante du réel, une région qu'il est possible d'en isoler suffisamment pour en déterminer le comportement". (Meyerson archives, A 408/60).
- 29 Bensaude-Vincent, 2001.
- 30 "Mais nul corps de la nature n'est de la matière proprement dite", Venel, article « Principes », in D'Alembert, Diderot, *Encyclopédie ou Dictionnaire raisonné des sciences, des arts et des métiers*, vol.3.

- 31 Bachelard, 1940, Engl. translation p. 45.
- 32 Duhem, 1902, chap. 1.
- 33 As early as the end of the 13th century, the *Summa perfectionis*, falsely ascribed to Geber, developed corpuscular views intertwined with the doctrine of principles. The pseudo-Geber described the combination of *minima partes* or “small particles”, which come together in a “strong composition” to form the two constituent principles of metals: sulphur and mercury. Newman William R., 1991.
- 34 Clericuzio, 1990, Principe L., 1998.
- 35 Bachelard, 1940, p. 46, 76. According to Bachelard, chemists work with three layers of reality: the “naïve” identification of particular substance; a “rationalized” and intellectual notion of substance; and the third layer of ‘non-substantialism’. Nordman convincingly argues that Whitehead and Latour later developed similar notions. (Nordman, 2006).
- 36 I use the ancient term “mixt” (derived from the Greek *mixis*, rather than the standard term “mixture” in order to avoid confusions with the modern notion of mixture, which is clearly distinguished from “combination”.
- 37 Aristotle, *De generatione et corruptione* I.10, 328a13f.
- 38 Needham, 1996, 2006.
- 39 Kuhn, Thomas, 1959, “The Essential Tension. Tradition and Innovation in Scientific Research”, in *The Essential Tension, Selected Studies in Scientific Tradition and Change*, Chicago, The University Press of Chicago, 1977.
- 40 Lavoisier, 1789, second section. “Chapitre sur le Tableau des Substances Simples”.
- 41 The compositional interpretation of the chemical revolution has been emphasized by Robert Siegfried and Betty J. Dobbs, 1968.
- 42 Guyton de Morveau et al. 1787, p. 14.
- 43 Condillac, 1780, II, VII, p. 413.
- 44 Duhem, 1902, p. 5.
- 45 Meyerson, 1931, §54–55, pp. 84–85.
- 46 Meyerson, 1931, §112, pp. 191–93.
- 47 Meyerson convincingly argues that Stahl’s doctrine of qualitative principles was more rational than Lavoisier’s theory because Stahl assumed the conservation not only of the quantity of matter but also of qualities such as combustibility. See Meyerson, 1921, appendix II “The resistance to Lavoisier’s Theory”.
- 48 Bohr, 1958.
- 49 Urbain, 1921, p. 11.
- 50 Charpentier-Morizé, 1997; see Bensaude-Vincent, 1999.
- 51 Simoes, Gavroglu, 2001. According to Roald Hoffmann some chemists deny the reality of orbitals to this day.
- 52 E. Meyerson, Conférence du 23 avril 1911 sur “Evolution de la pensée allemande dans le domaine de la philosophie des sciences”, p. 22. (Archives A 408/11).
- 53 Cassirer, 1910, chapter “Conceptualization in Natural Science”.
- 54 Bachelard, 1940, p. 57.
- 55 Simões, Gavroglu, 2001.
- 56 Wheland assumed the reality of molecules and double bonds, while he considered resonance a “man-made concept”, which described “the mental process of the person who makes the statement”. Pauling fundamentally disagreed. For him, classical structures, molecular orbitals and resonance were all equally man-made concepts. Simões, Gavroglu, 2001, p. 66.
- 57 Personal communication, September 05, 2006.
- 58 Klein (ed.), 2001.
- 59 Bachelard, 1940, p. 60.
- 60 Hacking, 1983.
- 61 Potential is the state of simple possibility in a real being. Interestingly, this notion seems to be reactivated by Jean-Marie Lehn in his program of dynamic combinatorial chemistry based on dynamically generated “virtual libraries” in which all possible combinations of the components are potentially accessible (Lehn, 1999).

- 62 I am grateful to Pierre-Michel Vauthelin for drawing my attention to this treatise.
- 63 Susan Linquist quoted by Zhang, S., 2003, p. 1177.
- 64 Drexler, 1995, p. 2.
- 65 Kuhn, 1977, p. 5. It strikes me that Thomas Kuhn, who acknowledged his debt to Meyerson, Metzger and Koyré, when he advocated these relevance of history for the philosophy of science, nevertheless rejected their philosophical views. While he praised Meyerson's and Brunschvicg's historical practices, and recommended them to his students, he dismissed their philosophical messages. Moreover, he never tried to explore the issues for which he reproached them, and he shaped his own view about the functions of history in philosophy without reference to the French tradition, *ibid.*, p. 11.
- 66 Canguilhem, 1977, p. 19.
- 67 From this perspective, the relevance of Aristotelian notions for chemistry is not to be viewed as a residue of the past that somehow survived the Scientific Revolution. Just as eighteenth-century chemists reinvented a four-element theory, today materials chemists are inventing an *analogon* of Aristotle's four-cause theory when they argue that the design of a new material is not a linear sequence running between function, property and structure. They rather take four parameters into account simultaneously; structure, property, performance and process. The Aristotelian component, like the Lavoisian component, is part of the "obligations" that help chemists to forge their own philosophical positions and their own system of values through confrontations with others.

Bibliography

- Bachelard Gaston (1940), *La philosophie du non*, Paris, PUF; *The Philosophy of No: A Philosophy of the Scientific Mind*, trans., New York, Orion Press, 1968.
- Bachelard Gaston (1952), *Le matérialisme rationnel*, Paris, PUF.
- Baird Davis, Scerri Eric and McIntyre Lee (2006) (eds.), *Philosophy of Chemistry. Synthesis of a New Discipline*, Dordrecht, Springer.
- Bensaude-Vincent Bernadette and Stengers Isabelle (1993), *Histoire de la chimie*, Paris, La découverte; *A History of Chemistry*, trans., Cambridge, Harvard University Press, 1996.
- Bensaude-Vincent Bernadette and Stengers Isabelle (1997), *Eloge du mixte. Matériaux nouveaux et philosophie ancienne*, Paris, Hachette Littératures.
- Bensaude-Vincent Bernadette and Stengers Isabelle (1999), "Atomism and Positivism: A Legend about French Chemistry", *Annals of Science*, 56, pp. 81–94.
- Bensaude-Vincent Bernadette and Stengers Isabelle (2001), "Graphic representations of the periodic System of Chemical Elements" in U. Klein (ed.), *Tools and Modes of Representation in the Laboratory Sciences*, Dordrecht, Boston, Kluwer Academic Publishers, pp. 133–161.
- Bensaude-Vincent Bernadette and Stengers Isabelle (2001b), "The Construction of a Discipline: Materials Science in the U.S.A", *Historical Studies in the Physical and Biological Sciences*, 31, 2, pp. 223–248.
- Bensaude-Vincent Bernadette and Stengers Isabelle (2003), "A Language to Order the Chaos", in M. Jo Nye (ed.), *The Cambridge History of Science*, vol. 5: *Modern Physical and Mathematical Sciences*, Cambridge, Cambridge University Press, pp. 174–190.
- Bensaude-Vincent Bernadette and Stengers Isabelle (2005), "Chemistry in the French tradition of Philosophy of Science: Duhem, Meyerson, Metzger and Bachelard", *Studies in History and Philosophy of Science*, 36, pp. 627–648.
- Berson Jerome A. (2003), *Chemical Discovery and the Logicians' Program. A Problematic Pairing*, Weinheim, Wiley-VCH.
- Bohr Niels (1958), *Atom Physics and Human Knowledge*, New York, Wiley.
- Bushan Nalini and Rosenfeld Stuart (2000) (eds.), *Of Minds and Molecules: New Philosophical Perspectives on Chemistry*, Oxford, Oxford University Press.

- Canguilhem Georges (1977), *Idéologie et rationalité*, Paris, Vrin.
- Cassirer Ernst (1910), *Substanzbegriff und Funktionbegriff: Untersuchungen über die Grundfragen der Erkenntniskritik*, Berlin, trans., *Substance and Function and Einstein's Theory of Relativity*, New York, Dover, 1953.
- Charpentier-Morizé Micheline (1997), *Perrin. Savant et homme politique*, Paris, Belin.
- Clericuzio Antonio (1990), "A Redefinition of Boyle's Chemistry and Corpuscular Philosophy", *Annals of Science*, 47, pp. 561–589.
- Condillac Etienne Bonnot de (1780), *La Logique ou l'art de raisonner*, Paris, Vrin 1981.
- Crosland Maurice (1962), *Historical Studies in the Language of Chemistry*, London, Heineman, 2nd ed., New York, 1978.
- Drexler Eric (1995), "Introduction to nanotechnology", in Markus Krummenacker and James Lewis (eds.), *Prospects in Nanotechnology. Proceedings of the 1st General Conference on Nanotechnology: Developments, Applications, and Opportunities, Nov. 11–14, 1992*, Palo-Alto, John Wiley.
- Duhem Pierre (1902), *Le mixte et la combinaison chimique*, Paris, Fayard, 1985; trans., P. Needham, *Mixture and Chemical Combination and Related Essays*, Dordrecht, Kluwer, 2002.
- Grosholz Emily R. and Hoffmann Roald (2000), "How Symbolic and Iconic Languages Bridge the Worlds of the Chemist: A Case Study from Contemporary Bioorganic Chemistry", in Nalini Bushan and Stuart Rosenfeld (eds.), *Of Minds and Molecules: New Philosophical Perspectives on Chemistry*, Oxford, Oxford University Press, pp. 230–247.
- Guyton de Morveau Louis-Bernard, Lavoisier A. L., Berthollet, C., Fourcroy (1787), *Méthode de nomenclature chimique*, Paris, Cuchet; reprint, Center for History of Chemistry, Philadelphia, 1987.
- Hacking Ian (1983), *Representing and Intervening*, Cambridge, Cambridge University Press.
- Hoffmann Roald (1995), *The Same and Not the Same*, New York, Columbia University Press.
- Hoffmann Roald (2005), "What Might Philosophy of Science Look Like if Chemists Built It", Boston University Colloquium for Philosophy of Science, October 20, 2005, forthcoming in *Synthese*.
- Holmes Frederic L. (1993), "La chimica nell'eta dei Lumi" in *Storia delle scienze Natura e Vita*, vol. 3, *Dall'Antichità all'illuminismo*, Torino, Giulio Einaudi, pp. 478–525.
- Klein Ursula (2001) (ed.), *Tools and Modes of Representation in the Laboratory Sciences*, Dordrecht, Kluwer.
- Kuhn Thomas (1959) "The Essential Tension. Tradition and Innovation in Scientific Research", in *The Essential Tension, Selected Studies in Scientific Tradition and Change*, Chicago, The University Press of Chicago, 1977.
- Lavoisier Antoine Laurent (1789), *Traité élémentaire de chimie*, Paris, Cuchet; trans., R. Kerr, *Elements of Chemistry*, Edinburgh, 1790.
- Lehn Jean-Marie (1999), "Dynamic Combinatorial Chemistry and Virtual Combinatorial Libraries", *Chemistry European Journal*, 5(9), pp. 2455–2463.
- Lévy Monique (1979), "Les relations entre chimie et physique et le problème de la réduction", *Epistemologia*, 2, pp. 337–370.
- Liegner Christoph and Guiseppe Del Re (1987), "The Relation of Chemistry to Other Fields of Science: Atomism, Reductionism, and Inversion of Reduction", *Epistemologia*, 10, pp. 269–283.
- Meyerson Émile (1921), *De l'explication dans les sciences*, 2 vols, Paris, Payot; trans., Mary-Alice and A. Siple, *Explanation in the Sciences*, Dordrecht, Kluwer, 1991.
- Meyerson Émile (1925), *La déduction relativiste* Paris, Payot; trans., *The Relativistic Deduction*, Dordrecht, Reidel, 1985.
- Meyerson Émile (1931), *Du cheminement de la pensée*, 3 vols, Paris, Alcan.
- Meyerson Émile (1933), *Réel et déterminisme dans la physique quantique*, Paris, Hermann.
- Needham Paul (1996), "Aristotelian Chemistry: A Prelude to Duhemian Metaphysics", *Studies in History and Philosophy of Science*, 26, pp. 251–269.
- Needham Paul (2000), "What is Water", *Analysis*, 60, pp. 13–21.
- Needham Paul (2006) "Aristotle's Theory of Chemical Reaction", see Baird, 2006, pp. 43–67.
- Newman William R. (1991), *The Summa Perfectionis of Pseudo-Geber*, Leiden, Brill.

- Newman William R. and Lawrence M. Principe (2002), *Alchemy Tried in the Fire*, Chicago, The University of Chicago Press.
- Nordmann Alfred (2006) "From Metaphysics to Metachemistry", see Baird, 2006, pp. 347–362.
- Nye Mary J. (1993), *From Chemical Philosophy to Theoretical Chemistry*, Berkeley, University of California Press.
- Primas Hans (1983), *Chemistry, Quantum Mechanics and Reductionism*, Berlin, Springer.
- Principe Lawrence M. (1998), *The Aspiring Adept. Robert Boyle and His Alchemical Quest*, Princeton, NJ, Princeton University Press.
- Scerri Eric (1991), "Electronic Configurations, Quantum Mechanics and Reduction", *British Journal for the Philosophy of Science*, 42, pp. 309–325.
- Scerri Eric (1994), "Has Chemistry Been at Least Approximately Reduced to Quantum Mechanics?", in D. Hull, M. Forbes and R. Burian (eds.), *Philosophy of Science Association*, 1, East Lansing, MI, pp. 160–170.
- Scerri Eric (2000), "The Failure of Reduction and How to Resist Disunity of the Sciences in the Context of Chemical Education", *Science & Education*, 9, pp. 405–425.
- Scerri Eric and McIntyre Lee (1997), "The Case for the Philosophy of Chemistry", *Synthese*, 111, pp. 213–232.
- Schatzki Theodore R., Knorr Cetina K., and Von Savigny E. (2001) (eds.), *The Practice Turn in Contemporary Philosophy*, London, Routledge.
- Schummer Joachim (1997), "Scientometric Studies on Chemistry II: Aims and Methods of Producing New Chemical Substances", *Scientometrics*, 39, pp. 125–140.
- Schummer Joachim (2006), "The Philosophy of Chemistry. From Infancy Toward Maturity", see Baird, 2006, pp. 19–39.
- Siegfried Robert, and Dobbs Betty J. (1968), "Composition: A Neglected Aspect of the Chemical Revolution", *Annals of Science*, 24, pp. 275–293.
- Simões Ana (2002), "Dirac's Claim and the Chemists", *Physics in Perspective*, 4, pp. 253–266.
- Simões Ana and Gavroglu Kostas (2001), "Issues in the History of Theoretical and Quantum Chemistry (1927–1960)", in Carsten Reinhardt (ed.), *Chemical Sciences in the 20th Century*, Weinheim, Wiley-VCH, pp. 51–74.
- Stengers Isabelle (2006), *La Vierge et le Neutrino. Les scientifiques dans la tourmente*, Paris, Seuil.
- Urbain Georges (1921), *Les disciplines d'une science, la chimie*, Paris, Doin.
- Zhang Shuguang (2003), "Fabrication of Novel Biomaterials Through Molecular Self-Assembly", *Nature Biotechnology*, 21(10), pp. 1171–1177.

Part IV
Life Sciences

Pharmacology as a Physical Object

François Dagognet

1 Introduction

It may seem strange for a “philosopher” – or at least someone considered as such – to devote attention to the notion of medication and what lies beyond it, the corresponding discipline of pharmacodynamics. This problem should concern the physician and none other, as he is responsible for prescribing and understanding those substances that are liable to cure. But already what complicates the examination and keeps philosophy away from this field of possible reflection is the fact that medication today has partly deserted medicine for chemical industry, which defines it, renews it and produces it.

Incidentally, the notion of medication can be practically substituted for a quasi-synonym, that of remedy. Yet the former refers to a chemical molecule, whereas the latter has a broader meaning, including among other therapies, balneotherapy, cure of fresh air, diet, etc. But the term of medication as it does not cease to develop came to overshadow that of remedy.

In what sense can medication attract the theorist of technology or a philosophy that has devoted only slight attention to curing?

We believe that there are several reasons for the philosopher to take part in a general reflection on therapeutics. First, any material (including chemical ones) should call for a theorized treatment. We make this maxim our own: “Philosophy is a reflection for which any foreign matter is good – and we hasten to add – for which any good matter should be foreign”.¹ Furthermore, medication has a privileged status, in the sense that it lies at the junction between physiology, the human that its task is to cure, and physico-chemistry, through a molecular renewal. One cannot find a richer or more contrasted realm, for it involves both extremes. Another reason for a requisite philosophical reflection arises from the fact that medication, both in

F. Dagognet (✉)
UFR de philosophie
Université Paris I Panthéon-Sorbonne
17 rue de la Sorbonne
75231 Paris Cedex 05, France

its constitution and its effects, borders on that which belongs directly to psychology. A case in point is “placebo” experimentation. A pill containing an active substance is given to some sick persons, whereas the same pill with respect to shape, color and size, but inert and devoid of any active substance is given to others. Now, it is not to be ruled out that the beneficiary may be more altered and struck in the second case than in the first; this has the effect of irrealizing the medication and obliges us to recognize in passing the importance of suggestion, which has an undeniable effect on the outcome. The false has a greater effect than the real. But the one who was carrying out the experiment knew that in principle one of the pills should not have worked. It had then to be recognized that this information seeped out and came to disturb the reading of the procedure. Henceforth, it will be necessary to institute a double blindfold: neither the one who is carrying out the experiment nor the subjects who are being tested on know where lies either the true or the false. It was necessary to go a step further and to conceal the very name, real or fictitious, designating the substance, because it may itself alter the result, as it is often a scientific name that impresses.

Furthermore, the experimenter has reached an implausible situation, one in which a false medication proves to be more active than the real one, causing faintness, vomiting and an unsteady step – a complete hysterized scene. The whole of pharmacology comes out shaken, as its object of study cannot be easily and surely circumscribed.

Manufacturers must nevertheless resort to the trick of the placebo, because before putting on the market a new medication, they are obliged to subtract the imaginary part from the active substance that it contains. They think they are retaining only the potentiality of the true; their analysis cannot remove the indistinct area surrounding the molecule responsible for the cure or at least the sedation.

Another reason for leading the philosopher into the realm of pharmacology is that the contraceptive pill (due to Pincus) has altered the very notion of medication: until then medication helped to prolong the duration of life and to hinder illness (eradication); henceforth, it no longer heals but merely prevents the consequences of sexual intercourse (childbirth). This led to a revolution in mores: the autonomy of the subject is enhanced; what the physiology of reproduction imposed is truly overcome. The reference to the philosophical is thus necessary, as the pill shakes the basis of society.

That is not all: pharmacology, the technique of curing, will quickly raise moral issues, which should not come as a surprise as we are touching both the basis of corporality and the foundations of the polis. We shall give a short and small sample. For the pharmacologist it is a question of evaluating the effects of a substance capable of curing; now, before making it available, one will have to test it in small quantities on a sick person. Obviously, in such a case one will interrupt the treatment that had been followed, for otherwise the discriminating test would be disturbed. But is it acceptable that we may and should stop the treatment on the pretext of pursuing a simplified reading? Are we not considering the person we are experimenting on as a “guinea pig” and without any benefit for them? We must have obtained the agree-

ment of the sick person, but such consent only goes to prove that we must have mystified them somewhat.

We await the answer of the moralist in the face of this conflict of duty: on the one hand, to encourage research with respect to the efficiency of the new treatment, on the other, to deprive of care the person whose health we worsen. The philosopher could recommend an eclectic solution – to lessen the risks, resorting to minute quantities, but this amounts less to resolving the problem than to avoiding it. All in all the notion of medication leads us to encounter problems of different orders: technical problems (those of manufacturing), epistemological ones (one must quit the realm of magic as well as that of empiricism), psychological ones (revealed by the placebo) and psycho-moral ones; the latter increasing as they are bound up with the development of this discipline.

These preliminary remarks lead us to realize that pharmacology concerns a narrow domain, especially when compared with that of mathematical analysis, astrophysics and microphysics. But precisely on account of its narrowness and its limits, as also because it pertains primarily to pragmatic questions of efficiency, pharmacology requests the attention of the philosopher, in that intelligibility or rationality find here their trial; no other discipline is more in danger of charlatanism or pseudo-naturalistic explanations. And when a theorist is searching for imaginary explanations in order to question them, he is sure to turn toward the books of apothecaries. An epistemological outlook is thus required to free the logic of remedy from an empiricism that it has yet to cast off.

2 The First Moment, the First System

Medical treatment, whose slow development we shall outline in order to emphasize it and first of all to acknowledge it (we foresee three periods), gradually acquired its independence, but it has come a long way. It began essentially with the Greeks (first and foremost Hippocrates and Galen) who adopted solutions without risk, almost irrefutable ones: first of all, the maxim “do no harm” (*primum non nocere*). To this effect the physician finds his inspiration explicitly in nature (*natura sola medicatrix*). How not to praise this caution when it refuses the strangest beverages alleged to be panacean?

Several consequences follow from this starting point: on a par with a Cosmos itself well-balanced, a healthy body owes its balance to the coexistence within it of humors antagonistic but neutralizing one another and which, each after their manner, reflect the very principles of the Universe (earth, air, fire, and water). These four operators, which mix and especially harmonize, are found in us: fire corresponds to yellow bile, earth to black bile – or cholera adust, water to lymph, air to blood. The quaternary is found, as expected, in the eventual predominance of one of these fluids accounting for the different temperaments (the sanguine, the melancholic, the phlegmatic, and the choleric).

Under these conditions the disciple of Aesculapius finds the way paved for him: his aim is to reestablish the harmony endangered by the illness; he will endeavor to get rid of the excess or to augment the depleted. Sometimes, if he is unable to establish concord – the morbid becoming encysted or abscessed – he will resort to a surgical solution – the opening of a canal in order to expel the humor that was solidifying.

Hippocrates was to allot an important place to diet. And Plato in the *Republic* will take up this lesson; aside from nourishment, the philosopher was to recommend exercise, gymnastics or again baths or a “cure of fresh air” – all that activates and prevents stoppage (deposits, flatulences, catarrhs), all that avoids the “crasis”:

When his heroes are on campaign he does not feast them on fish, although they are on the shore of the Hellespont, nor on boiled meat, but only roast. If that's your view I assume that you don't approve of the luxury of Syracusan and Sicilian cooking.²

Indeed, Plato takes up the thesis that it is essential to avoid food that gives somnolence as well as repletion. According to the case at hand, a tonic, astringent or laxative nourishment will be advised.

Hippocratic medicine does not hesitate to seek in the plants that surround us the means to reinforce our defenses.

Whatever the prescription, it is essential to be attentive to humoral balance: that is why “The supposed delights of Attic confectionery” or even seasonings as well as the slightest variety are banned, because they all disturb “the wisdom of the body”; likewise it is essential, with respect to plants, to keep to simples.

Galen was to renew the problem of therapeutics. He devotes much space to the excipient: although inert it enables the medication to have an effect, for it is less a question of “ordering” it than of preparing it and choosing the form that will allow it to be introduced in the most appropriate way (intradermic or digestive). With the help of the mortar in which drugs are pounded, Galen and his disciples present us with pomades, ointments, plasters, lotions, powders, and pellets. Moreover, Galen was to enlarge the list of preparations: fumigations and troche.

Pharmacology is not independent from that which commands it – to not exceed the body, but to free it merely from what intrudes it, to fight against stasis or flux. But the theory of Hippocratic inspiration did not fail to recognize in the meaning of the term *pharmakon* both a possible remedy and a poison. This is a way of recalling that one must act according to nature and not do violence to it.

During the Middle Ages and until the 18th century pharmacology was content to continue and enhance the dominant naturalist doctrine; we shall nevertheless note some advances, the emergence of the better but still accompanied by the worst.

First, pharmacology will follow the so called “theory of signatures”, a hermetic principle; a plant or an animal would be chosen to the extent that they resemble the sick person or their pathology in some aspects with respect to their shape, color or size. A narcotic such as mandrake achieved its central place because its root brought to mind the human body; one could even discern in its forked shape two arms and two legs. Yellow-flowered gentian, recommends itself against hepatic affections (icterus jaundice). However – and this is a small innovation in the system – the 18th century added metals to the admissible ingredients of medication: gold foremost,

because inalterable, gold dissolved beforehand or *aurum potabile*. Others follow (mercury, quicksilver). One will also resort to animal organs: vipers provide the best-known and prestigious compound, theriac, which comprises sixty elements (a sort of panacea or universal remedy). Why this reptile in particular, if only because it renews continuously its skin – a sign of strong vitality? This improbable doctrine is used to justify anything.

Let us add that bloodletting is often recommended for the same reasons – to fight against any extravasation (apoplexy).

Yet the system can also claim several unquestionable results that will strengthen it. At the end of the 18th century for instance, seafarers return with tropical plants that deserve to receive a special status, because the pharmacologist will extract “specifics”, – quinine and soon after ouabain, strychnine, pilocarpine papaverine, etc. We are beginning to leave behind the realm of vesania, but also to glimpse what will come – a medication that is to lose its aura of magic and strangeness, although the newcomers still keep up the spirit of the past.

Among the modestly positive side that gives credit to the naturalist theory and prevents its demise (this has not yet occurred, as the theory continues to be in use here and there), we may mention a discovery made in an indefensible manner, because one arrives at the true through the false.

The therapist or even the mere amateur might have observed near rivers or streams flourishing willows. Now, the physician holds water to be harmful, whether it runs out from the tissues or spreads through them. He is led to seek what allows the willow to defend itself and can only point to the bark – the part that was exposed and should therefore have been damaged. In due course a glucoside was extracted that was called salicin (from the Latin *salix*, the willow). It was quickly transformed into salicylic acid, and (in 1853) into acetylsalicylic acid – our aspirin, which is efficient against inflammation, consecutive pain and articular swelling – a pathological state we designate metaphorically as hydric. Let us add that those affected by this state are advised to live and move around in a dry atmosphere; they dread fog and whatever is humid. Through such indirect and false routes, we discovered the power of *Digitalis purpurea* (observed by William Withering in 1785) which prevents cardiac insufficiency; the French pharmacist Claude Nativelle purified this digitalin (1844). *Colchicum*, which is effective against gout, was discovered in a similarly indirect manner. Hippocratic doctrine does not lack support and effects, which explains its persistence; it continues to influence surreptitiously therapeutics, blurring the notion of medication. At the end of the 18th century two major innovations will revive Hippocratic theory, bringing it to its height. Vaccination is brought back from Constantinople to Europe by the wife of the British ambassador to Turkey; she had observed how the Asians protected themselves against smallpox, that is by inoculating under the skin a small pustule of the illness – as if the lesser makes it possible to hinder the greater. Later Jenner improved this defensive gesture. Another comparable strategy: Hahnemann, after having taken quinquina (the bark from this exotic tree) which produces hyperthermia, observed a sudden rise of temperature; he was to draw as early as 1796 the so-called law of similitude – the same cures the same, under the condition that one resorts to minute doses. Mithridatization takes on new colors. Indeed, King Mithridates, in order to

avoid an eventual poisoning, was used to take the most violent poisons in gradually increased doses without been otherwise indisposed. Nature it seems would have provided us with the major weapon against pain and illness – that is tolerance.

We continue to note as concerns therapeutics a mixture of imaginary and real. *Materia medicans* continues to perplex us, because the effective appears often in the midst of the strange.

3 The Second Stage

The second stage of our broad overview begins in the mid-19th century and ends around 1950. It is a perplexing stage. For instance, we must acknowledge that Claude Bernard and those associated with him – Magendie as well as Vulpian – altered completely the notion of medication (as we shall see), but did not discover any of importance. Why then give them a special place? It would be difficult to find a text more novel than *Leçons sur les effets des substances toxiques et médicamenteuses* (1857), actually more specifically devoted to the analysis of poisons than remedies. But Claude Bernard was to set himself primarily the task of “desubstantializing” what had been overly reified. Indeed, Claude Bernard showed that medication gives different results according to the animal on which it is tested, the dose given, the manner of administration, the preparation, the illness, the duration of prescription, the association with other substances (hence what would be called the potentialization, for two small doses combined are stronger than what a mere addition might lead to expect), the time of taking and even the means by which the results are detected, all which has the consequence of relativizing the evaluation. Claude Bernard does not go so far as to lapse into uncertainty, but he removes medication from the realm of empiricism.

It is constantly repeated that hypotheses must be verified; anything novel – the idea – must be put to the test of reality, for the true is the verifiable, but we must point out that according to our parameters we reach questionable results or conclusions. This polyvalence, which is troubling, casts doubt on the experimental set-up. It is necessary to call on other methodologies.

We are easily surprised. For instance:

If one ingests mercury cyanide into the stomach of a healthy dog and into that of a sick dog, the healthy dog dies immediately whereas the sick one dies very slowly [...]. Mercury cyanide killed the healthy dog because of the free cyanohydric acid that was produced in presence of gastric juice. The sick dog did not succumb by a different mechanism; the slowness of its death however should be ascribed to the fact that the mercury cyanide did not find in the sick stomach the gastric juice from which was freed the cyanohydric acid.³

Another disappointment similar to the latter: we thought we could hold as certain the thesis that “physiology, then pathology and finally pharmacodynamy” promote one another; their inseparability being truly at the base of experimental medicine. Indeed, from this unitary point of view, pathology corresponds to an impeded physiology (to a greater or lesser degree – the difference being merely quantitative), but

this illness is only well observed through what modifies it. If I am able to slow down an excessive secretion by chemical means, I understand better the clinical picture; the suppression of the cause must result in that of its effects. Medication would play an explanatory and differentiating role. We have mentioned Claude Bernard's theory favoring unity; yet he neither follows it nor applies it.

Proof or illustration? Claude Bernard put forth the central and revolutionary notion of "internal environment or *milieu intérieur*", two words which appear to contradict one another and yet designate life in its capacity to escape external determinants. Claude Bernard will search for all the constants of the organism, namely the blood sugar constant, in other words the "one gram of glucose" per liter of blood. Following the preceding remarks, illness is conceived as a violation of the equilibrium (diabetes). But the patient will simply be advised to ensure by nutrition what he no longer controls; he will have to avoid both the lack and the excess of sugar.

The most worthy of note is that Claude Bernard discovers a physiology of importance, likewise he begins by explaining its mechanism, without however this advance being accompanied by any therapy. Later, during the 20th century, the discovery of insulin – thus named because secreted by the islets of Langerhans in the pancreas (a gland producing internal secretion, again two terms which contradict each other and for the same reason) will correct what the liver is accused of.

We shall bare in mind that Claude Bernard was unable to keep together the three moments, as he had recommended. For the first two – physiology and pathology – advance together, whereas the third did not benefit from such agreement; he even goes back to the notion of a medication of equilibrium – neither the too much nor the too little of a recommendation tinted with Hippocratism.

Why then praise Claude Bernard, whom we place at the center of an upheaval in therapy? First, he brought to an end a long tradition, difficult to dispel, one that made every effort to "chase away the disorder, by freeing the body that had been invaded by it", whence the use of the cupping glass, the purge, the application of leeches, etc. One recommends also mountain air which is reinvigorating.

But the end of the 19th century will witness an unprecedented renewal of pharmacology; we find ourselves at a turning point, in spite of some remains of the past.

First, we shall present briefly the most discussed case. Claude Bernard learned that the Indians of the Amazon mixed together what they had obtained from a kind of vine named *curare*, which they applied to their arrowheads. By this means, they paralyzed quickly the animal, whose mobility returned after a short lapse of time. Briefly put, we have here the case of a transitory intoxication particularly valued by hunters (an easy capture). Claude Bernard did not fail to be struck and sought to understand the mechanism of this feat. He went on to observe or rather experiment that the poison affects neither the nerve of irritability nor that of reaction or contractility.

Already Haller had been led to distinguish sensibility from motor functions. Claude Bernard would suggest that the poison could operate only at the time of

passing from one to the other of these two nerves in the “motor plate”. Physiology is beginning to understand the power and role of junction and mediation.

Another Bernardian discovery of importance will reinforce the former, even if it does not lead either to the elaboration of a particular medication; it nonetheless alters medicine and brings to light a hitherto unknown territory.

When the physiologist stimulates the nerve of the tympanum cord, which goes to the submaxillary gland, this provokes a distinct over-activity of the capillar blood vessels, such is the effect of the nervous system (the object of slight attention here) on circulation and all that depends on it. Gradually comes to light the existence of local circulations (a departure from Harvey’s general circulation) and especially that of a twofold system – the sympathetic and the parasympathetic, a constrictor and a dilator (sometimes the reverse according to regions and functions).

Let us retain this lesson: it is less important to pay heed to the organs and their physiology strictly speaking than to that which higher up insures their regulation – the stimulants and reducers. What commands prevails over what performs. The outcome, soon to follow, will be a harvest of medications concerning centers, including the brain; the whole pharmacopoeia is concerned with what allows or inhibits junction (the antisynapses, chemical mediators).

With Claude Bernard, the pioneer, began a change, which in the early 20th century witnesses an acceleration. We see nothing more decisive than what O. Loewi was to bring to light. A frog’s heart had been separated from any connection with its nerves, in particular the pneumogastric nerve (still called the “vagus nerve”, because it gives rise to variable if not surprising effects); it is nevertheless irrigated by a liquid that bathed another live heart; the latter has been even stimulated causing its slackening. But if the nerve of this heart has had to lessen its rhythm, it is also going to act (indirectly) on the former, deprived of any connection. How to explain this extension or the passage between two separate organs, if not by the fact that the live heart freed a soluble and fugitive substance, which will soon be identified (acetylcholine)? The most important conclusion does not reside where one would expect (one can modify at a distance the movement of the heart) but in the evidence of a possible influence of one organ on another (the same) by means of a “chemical mediator”. Henceforth, pharmacology will be capable of acting anywhere in the organism, either with the aim to moderate such and such a function, or with the intention of increasing it (later psychotropics will act on conduct or behavior). Thus is confirmed the general theory of transmissions or exchanges – the very basis of physiology and pathology.

4 The Modern Stage

We do not claim to present the composition of the major medications – the list could be undoubtedly narrowed owing to the fact that many drugs are simply copies of the primary ones – nor to examine their properties – a task of compilation. We merely propose to bring out one of the most distinctive achievements of the discipline (the

renewal takes place in 1950). We leave behind Bernard's doctrine and his methodology as presented in *An Introduction to the Study of Experimental Medicine*, because pharmacology imposes the change of status and attitude both in therapy and physiopathology.

The philosopher cannot remain insensitive to this change; he must go beyond the exercise he was used to practicing and recommending – the coupling of the idea with the real that could validate it.

This neo-pharmacology was from the outset confronted with a problem that it was unable to solve: modified cells, parasites barricade themselves, protect themselves, to speak by images; actually, the organism defends itself against them and ends up isolating them. The mediation then fails, because it did not penetrate into the metabolism of its enemy. The therapist can increase the dose as much as he wants, he does not necessarily solve the problem – to cure the sick person, whom he may even weaken. That is why pharmacology must henceforth pay heed primarily to the strategy that will allow it to break or to bypass the protecting scleroses of the invader as well as its plasma membrane. What matters is to have access to the sanctuarized illness, but how? And the physician remains powerless in front of an inaccessible opponent.

We should perceive remains of empiricism in this battle, as early solutions or at least their beginnings result from an unintentional loan from chemical industry. The chemistry of I.G. Farben in Germany and that of Rhône Poulenc in France led to recognize the power of certain molecules (methylene blue or prontosil rubrum) to break into the enemy fortress (the cell or the tissual fiber); once in the stronghold, they establish themselves, eventually destroying their hosts; the staining molecules succeeded the preliminary step, that of their entrance, owing to the fact of a kinship which will ultimately be established through laboratory research. The first sulphonamide drugs were to follow (namely prontosil).

Antibiotherapy follows the same pattern. Fleming noticed in 1928 that certain bacteria – staphylococci – were invaded by a woolly mold greenish white in color – a fungus of the genus *Penicillium notatum*; it denatured the microbial colony, preventing it from proliferating. Simultaneously a hen house was decimated by a comparable parasite of the genus *Streptomyces*.

In both cases, a fungus and what it secretes would gain entrance into the cells, causing them to perish.

Let us continue to ascend and to generalize: be it a stain or a parasite that develops at the expense of what it lives on, the success comes from bringing about the first phase of the operation – assimilation.

It is time to go beyond this factual movement and to reach a more general theory: why not systematically provide the pathological with that which it cannot refuse (a sure entrance) – what enhances it and favors its expansion, yet with a slight difference in its molecular structure. Nothing is more easily admitted than “the same”, but an “almost nothing” is added to it, and all the more minute as it is important not to complicate penetration; this minimum nonetheless acts on the enzymatic metabolism of the intruder; the enemy has fallen into the trap of an imperceptible difference but one sufficient to disrupt it. In other cases, one resorts to the reservoir of the body,

be it that of an ill person or a healthy one; one borrows a substance it diffuses in the body, modifying slightly its structure. The medication needs only to fulfill two conditions: an easy entrance into the opponent, but the slightest difference in its constitution with respect to a substantial to reduce or increase.

Pharmacology will go a little further, which diminishes its portion of simple fact: to seek to understand the mechanism which accounts for bacteriolysis; we already know that it is due to the inhibition of the synthesis of what will constitute the plasma membrane of the microbe (the former softens); and this morphological deficiency is the result of structural analogy, of a competition relative to the occupation of the place of “murexine”, itself involved in the protecting membrane of the bacteria. Several reactions are to be expected in that which penicillin invades, but the growth and subsistence of the microbial parasite are undermined. Let us enumerate a whole wealth of penicillins (pheneticillin, propicillin, azidocillin, hetacillin, carindacillin, piperacillin, etc.) all different in their range of action and with respect to their destructive capacities.

In the same vein we know that “chemical mediators” play a role in the organism: they insure the connections. Acetylcholine, as we noted, decreases the rhythm of the heart but is quickly destroyed by cholinesterase, because a power that endures is not tolerated. The slowed down heart soon resumes its regular beat; a regulator is not a break. Hence comes into existence a whole family of related medications: to prevent that which prevents, to inhibit that which inhibits, which leads to another kind of effectiveness.

We even believe to glimpse the future: when our “slightly different” compromises a function, what occurs exactly at the infra-cellular level? Is it a transmission of an electronic nature? We shall in the future come to understand better the nature of the mechanism of communication, for until now we know better the “what is” than its “how”.

One should be careful not to confuse this pharmacodynamy with homeopathic theory (presented by Hahnemann in his *Organon of the Medical Art of Curing*, 1811). According to this doctrine one must fight the disease itself by minute doses. It is not difficult to recognize a difference with pharmacodynamy based on a slight molecular modification.

Later arose a different philosophy of medication, diametrically opposed to the one we presented above: (a) Must one not experiment, and more so today than in the past, in order to be sure of the non-toxicity of a remedy (which both cures and poisons)? Is this not a return to Claude Bernard and his methodology? – (b) Does not the discovery of one or another antibacterial medication, with Fleming at the outset, prove that nature provides us with that which will save us? Are we not reverting to Hippocrates’ theory? Emphasis is placed similarly on the benefits of plant extracts coming from faraway lands: cocaine (coca leaves), quinine (cinchona bark), morphine (from poppy), emetine (from ipecac root), etc. Mystification resorts to plants as well as animals (organotherapy).

Yet we believed to have brought into question the soundness of these two pillars of pharmacology – Hippocratism and Bernardism – because in the final analysis the

former does not cure (setting value primarily on diet), whereas the other limits itself to physiology, having indeed not recommended any truly active medication.

On the contrary, current pharmacology has provided us with two quasi-panaceas, antibiotics and psychotropics. We can henceforth fight against the worst disorders that afflict us, both invasive illnesses (tuberculosis, meningitis) as well as so called illnesses of the soul (antidepressants and anxiolytics). Such a feat modifies the life of humankind; we shall soon be able to postpone the hour of death; we shall also get rid of the aftereffects of illnesses not completely stamped out or secretly subsisting.

It is not conceivable that the “philosopher” should remain absent from such a victory, all the more so that it carries with it both socio-economic and ethical problems.

5 Conclusion

We return to our initial remark: can philosophy, without relinquishing its status and role, concern itself with medication and endeavor to reflect on it? Does this task not belong solely to the therapist, the one whose duty is to enter into the mysteries of *materia medicans*?

We shall not stop at this question-objection: one could as well argue that the philosopher should not take part in discussions about medicine (because he is not a physician), nor again those relating to physics (as he is not a physicist). One gains nothing, on the contrary, by limiting the scope of philosophy.

First, we believe that any object – and *a fortiori* medication – deserves to be taken into consideration by the philosopher, because it always carries with it ingenuity and moreover cannot be separated from society which it expresses in its own way. We hold it for a “total social fact”; interiority, in our view, asserts itself and reveals itself in exteriority and its constructions; it is useless, fruitless to turn towards the self, where one acquires only “a series of rash assertions” as Nietzsche warns us (“What gives me the right to speak of an ‘I’, and even of an ‘I’ as cause, and finally of an ‘I’ as cause of thought?”).⁴ Nietzsche goes even further, for consciousness and the notion of subject derive from grammar; language would have imposed them on us from the start). In consequence, the most modest utensil, the least elaborate tool or the most heterogeneous materials deserve to be recognized, analyzed, rehabilitated. In this respect medication will not leave us indifferent, for it links the suffering man with his productions (the magistral preparation or the manufactured one). Claude Bernard claims that “the physician cures when he can, relieves when he cannot cure, consoles when he can no longer cure or relieve”. The remedy proves to be capable of this threefold operation, – objectivity coming to the aid of the most dramatic of intersubjectivities.

Beyond these considerations, pharmacology should mobilize the philosopher, because he is confronted here with a question of a trans-physical nature: how is an appropriate matter able to get around and to overcome a malignant cell or a dangerous

parasite, both guarding from penetrating into them what could destroy them? If the physician contented himself with breaking this protecting fortress by a direct attack (with the help of a fibrolytic or a fluidifier), he would open the extent of the body to the enemy. It goes without saying that by deceiving this enemy, which he favors at first, he quickly resorts to that which will suppress it. The simple strategy consisted in calling on two medications that are opposed but which combine and condition one another.

Pharmacology allows us to understand the success of obliqueness, that of tactical procedures. Pharmacology well understood goes beyond pharmacology and begins to mix with metaphysics, because it discovers in matter that by which to surpass it or at least to modify it.

Endnotes

- 1 Georges Canguilhem, 1966, p. 7.
- 2 Plato, *Republic*, Book 3, 404d. English trans. Desmond Lee, London, Penguin, 2007.
- 3 Claude Bernard, 1857, p. 104.
- 4 Friedrich Nietzsche, *Beyond Good and Evil*, § 16. English trans. R.J. Hollingdale, London, Penguin, 1990.

Bibliography

- Beaune Jean-Claude (1993), *La philosophie du remède*, Paris, Champ Vallon.
- Bernard Claude (1857), *Leçons sur les effets des substances toxiques et médicamenteuses*, Paris, PUF.
- Bovet Daniel (1988), *Une chimie qui guérit*, Paris, Payot.
- Canguilhem Georges (1966), *Le normal et le pathologique*, Paris, PUF, 1996.
- Canguilhem Georges (1989), *Écrits sur la médecine*, Paris, PUF.
- Chen Wāi (1992), “Sir Almroth Wright’s vaccine programme and the construction of penicillin” in *The Laboratory Revolution in Medicine*, Andrew Cunningham and Perry William (eds.), Cambridge, Cambridge University Press.
- Chast François (1995), *Histoire contemporaine des médicaments*, Paris, La Découverte.
- Griscelli Claude, Graillot-Pak Claude and Payen Geneviève (1993), (eds.), *Stratégie des thérapeutiques innovantes*, Paris, Éditions FrisonRoche.
- Lecourt Dominique (2004), *Dictionnaire de la pensée médicale*, Paris, PUF.
- Pignarre Philippe (1997), *Qu’est-ce qu’un médicament ?*, Paris, La Découverte.
- Pignarre Philippe (1990), *Ces drôles de médicaments*, Paris, Les Empêcheurs de penser en rond.

Philosophy of Biology: An Historico-Critical Characterization

Jean Gayon

1 Historical Prelude: “Philosophy of Biology”

Literally speaking, “Philosophy of biology” is a rather old expression. William Whewell coined it in 1840, at the very time he introduced the expression “philosophy of science”. Whewell was fond of creating neologisms, like Auguste Comte, his French counterpart in the field of the philosophical reflection about science. Historians of science know that a few years earlier, in 1834, Whewell had generated a small scandal when he proposed the word “scientist” as a general term by which “the students of the knowledge of the material world” could describe themselves, and distinguish themselves from artists. The term “philosopher”, Whewell argued, was too wide. A new generic term, more or less equivalent to the French term “savant”, was needed in order to prevent the disintegration of science that seemed to flow from its specialization in modern times.¹ When Whewell first introduced the terms “philosophy of science” and “philosophy of biology” in his 1840 *Philosophy of the inductive sciences*, the latter term was merely a special branch of “philosophy of science”. The expression “philosophy of science” itself had two justifications: firstly, this phrase expressed the idea that “science” remained cognitively coherent enough to justify a critical enquiry into its methodological unity and its foundation; secondly, the phrase “philosophy of science” was required in order to distinguish a properly “philosophical” enquiry from a “historical” approach to science. Although Whewell’s 1840 *Philosophy of the Inductive Sciences*² had approximately the same chapter structure as his 1837 *History of the Inductive Sciences*³ (that is, a series of chapters successively devoted to the concept of science in general, and although there was considerable overlap between the contents of the two books, then to particular sciences), its theoretical purpose was different. Clearly, Whewell was not

J. Gayon (✉)
Institut d’histoire et de philosophie
des sciences et des techniques
Université Paris I-Panthéon Sorbonne
13, rue du Four
75006 Paris, France
E-mail: gayon@noos.fr

willing to confuse the genres of history and philosophy as Auguste Comte had done in his 1830 *Cours de philosophie positive*.⁴ Note also that the word “biology” was still extremely rare in English when Whewell used it in 1840. In fact, Whewell’s 1837 *History of the Inductive Sciences* does not make use of the word “biology”: Whewell successively examines “botany”, “zoology”, “physiology” and “comparative anatomy” as special branches of “analytico-classificatory science”, then discusses palaeontology as a special case of the “palaeo-etiological sciences”. Three years later, in his *Philosophy of the Inductive Sciences*, Whewell does use “biology” as a generic term for all the sciences dealing with life. The various sciences which were separately examined in the previous book are now collectively considered. Furthermore, the main philosophical problem raised by biology is its dual nature: biology is both nomological and a historical science. Modern philosophers of biology are generally unaware of the story of the origins of the expression “philosophy of biology”, but Whewell’s dual theoretical nature of biology is still a major concern for modern “philosophy of biology”.

Let me go a little further in my historical prelude to “philosophy of biology”. Before Whewell, Auguste Comte had extensively used the apparently similar expression “biological philosophy” [*philosophie biologique*]. The whole series of lectures of the *Cours de philosophie positive* published in 1837 were devoted to “biology” (*lessons 40–45*). Historians of science traditionally agree that Comte made the word “biology” popular among scientists all over Europe. In the mid-1830s, it was a rare word, with no strong conceptual or institutional impact. Comte’s lectures on “biological science” not only advocated a unified appraisal of life sciences as a whole, they also included a detailed exposition of what Comte named “biological philosophy”. By this phrase, Comte meant the most theoretical part of biology, or the “fundamental conceptions” of biological science. This was in agreement with Comte’s method, which consisted of analyzing the fundamental conceptions of each theoretical science, one after another. Up to a certain point, Whewell agreed with this method. He indeed wrote repeatedly in his 1840 book that the philosophical enquiry into a special science consisted of extracting its “fundamental ideas”. However there is a dissimilarity between Comte’s “biological philosophy” and Whewell’s “philosophy of biology”. Comte used the expression “biological philosophy” exactly in the same way as contemporary authors used the old expression “natural philosophy”. Just as “natural philosophy” was a genuine science – the science that 19th century authors progressively came to name “physics” –, “biological philosophy” was by itself part of the “biological science”. It was its theoretical part. On the other hand, Whewell did not think that “philosophy of biology” was part of the biological science. It was rather a critical examination of its fundamental conceptions with the help of philosophical tools, “philosophy” then being something different from science.

After Whewell, the phrase “philosophy of biology” was occasionally used by some English or American authors (and by them as a keyword), but it remained rare till the end of the 1960s.⁵ In contrast, the expression “biological philosophy” was commonly used in the 19th and 20th centuries both in Continental and Anglophone countries, for all sorts of philosophical reflections on the life sciences, ranging from scientists’ or physicians’ commentaries on the fundamental conceptions of their discipline (e.g. Kurt Goldstein), to philosophers of science interested in biology (e.g.

Georges Canguilhem, Marjorie Grene), and general philosophers willing to think of life in general (e.g. Hans Jonas).

In the 1970s the expression “philosophy of biology” was massively adopted by philosophers of science in Northern America and other English speaking countries. This was probably the consequence of a famous paper published by David Hull in 1969 with the title “What Philosophy of Biology is not”.⁶ In this paper, the young David Hull criticized the way neo-positivist philosophers of science treated biology, that is, as an example among many of the ordinary methods of science in general, and of its basic theoretical unity. Instead, Hull advocated the idea of a specific philosophy *of* biological science, which could not be reduced to an application or exemplification of the standard doctrines of “philosophy of science” about explanation, theories, and other similar subjects. Although not all philosophers of science interested in biology endorsed Hull’s revolt against logical positivism and reductionism,⁷ in America they all adopted the expression “philosophy of biology” in a very short period of time. Within 3 or 4 years, “philosophy of biology” became a generic name for any kind of philosophical reflection on biology in America. In Europe, the expression diffused later, mainly in the 1990s. Today, it is widely used on an international scale, though the general philosophical community in continental Europe barely understands what the conventional expression “philosophy of biology” connotes, that is something different from traditional “biological philosophy”. In practice, few European philosophers are involved in the modern industry of “philosophy of biology”. Perhaps the country most open to philosophy of biology has been the Netherlands. England, France, Germany were late to join the game.

My historical overview stops here. The linguistic story that I have summarized may seem unimportant: why bother about a possible nuance between “philosophy of biology” and “biological philosophy”, or other expressions such as “epistemology of the life sciences”? Isn’t it just a question of words? Yes, it is a question of words. From a logical point of view, modern philosophy of biology as it has developed (mainly in Northern America) could have been named differently. For instance it could have been named “biological philosophy”, with no significant loss of content, in terms of the problems and solutions that have been explored by this special branch of philosophy of science. But this did not happen. In fact, the historical sketch that I have outlined suggests that the expression “philosophy of biology” had deep historical roots in the English-speaking philosophical world, and emerged as a conventional expression in a rather special context in the 1960s and 1970s. I will return to this problem in the general conclusion of this paper. Suffice to say that in this paper I will be concerned with a problem of cognitive style.

I will propose an evaluation of the distinctive features of modern “philosophy of biology”, as opposed to other traditions in philosophy of science that have dealt with the life sciences. I will first compare modern “philosophy of biology” with the general evolution of philosophy of science over the past 40 years. I will then try and characterize the difference between the “philosophy of biology” that emerged in Northern America in the last third of the 20th century, and other traditions that often designate themselves as “epistemology of the life sciences”. After this, I will take the journal *Biology and Philosophy* – the only journal specifically devoted to philosophy of biology (in any sense of the word) – as a test case for my analysis.

2 Philosophy of Biology in Relation to the Recent Evolution of Philosophy of Science

In the past 40 years, analytically oriented philosophy of science has become more open to history. At the same time, philosophers of science have expressed an increasing interest for new scientific areas and particular modern scientific theories. These two shifts went in the same direction: they testify to a certain skepticism regarding the idea of a general and timeless theory of science. This is sometimes referred to as the “historical turn”.

Closely related to the “historical turn” is the “regionalist turn”. In its most radical form the regionalist turn in philosophy of science can be summarized in the following assertion: nothing of interest in philosophy of science can be done outside the realm of within-discipline work. This is, of course, another way of rejecting the idea of “philosophy of science” in the strong sense of a general theory of science that would apply to any domain past, present or future.

It is worth locating modern philosophy of biology relative to these two general trends in modern philosophy of science.

- (1) Philosophy of biology has been a major illustration of the regionalist turn. Some philosophers of biology may well think that there is room for a general, discipline-independent theory of science, others may not. But all philosophers of biology are convinced that biology, especially biology as it exists today, requires a specific approach. Modern philosophers of biology, whether or not convinced of the autonomy of biological sciences, agree on the idea of the heuristic fertility of an autonomous development of philosophy of biology as a discipline or a sub-discipline.
- (2) Conversely, philosophy of biology has clearly not been concerned by the historical turn. The opposite is the case. The most obvious characteristic of philosophy of biology is the common acceptance of the methodological opposition between an analytical (or conceptual) approach to biology and an historical approach to biology. One should not confuse here “philosophy of biology” and “philosophers of biology”. In practice, a significant number of philosophers of biology are both philosophers and historians of biology. For instance, Michael Ruse or David Hull, who were so prominent in the construction of philosophy of biology as a discipline, have each made significant historical contributions. It is even probably true that they have written more in the genre of history of science (or history of philosophy) than in the genre of philosophy of science. But they have imposed a certain style of research based on the aforesaid distinction. Those who have explicitly challenged this view, and together claimed to be doing “philosophy of biology” are quite rare. Marjorie Grene and Richard Burian are probably the best examples. I am also one, though I come from a quite different cultural horizon. The non-historical character of philosophy of biology is in fact an implicit norm of behavior in the academic context, not a well-articulated doctrine. It constitutes a major difference with other traditions of reflection on biology among philosophers of science.

- (3) Philosophy of biology has been overwhelmingly concerned with modern biological theories, concepts and theories.

This is a direct consequence of the separation between history and philosophy of biological science. However, this does not mean that philosophers of biology have generally been open to scientific novelty. Most of the literature over the past 30 years has dealt with a relatively narrow spectrum of biological knowledge: evolution, classification, genetics. This, in turn, is another difference with other traditions.

3 “Philosophy of Biology” Versus “Epistemology of the Life Sciences”

Since it grew up in the typical manner of the American specialization and professionalization of academic research, philosophy of biology is easy to locate and characterize: it consists of journals, programs of research and education, chairs. In other words, it is a discipline in the institutional sense of the word. The intellectual traditions that have been or are in competition with “philosophy of biology” are not as easy to characterize, because their intellectual identity is not as clearly founded on internal disciplinary frontiers. The commonest expressions for these different traditions are: “epistemology of the life sciences” and “philosophy of the life sciences”. The first expression refers to “epistemology” as it has been commonly used in Continental Europe and Spanish-speaking countries throughout the 20th century. Thus it is not “epistemology” in the sense that the term is used in English or American philosophy, that is, theory of the foundations and limits of knowledge in general. In the continental tradition, “epistemology” is ordinarily taken as a synonym for “philosophy of science”, or, more precisely, philosophy of the sciences as they historically existed and exist (theory of given knowledge, rather than theory of what knowledge ought to be).⁸ In Germany and France, this is commonly called “historical epistemology”.

Let me try to give a positive characterization of “philosophy of biology” and “epistemology of the life sciences” as two styles in philosophy of science. By “positive characterization”, I mean a characterization in positive terms, not through rejection of another attitude.

The dominant attitude in epistemology of the life sciences (that is the view held by most European philosophers of science working on life sciences) consists of favoring the traditional historico-critical method. It is normally based on two deep convictions:

- (1) The sciences pose their own norms of rationality in the course of history. Accordingly, the role of philosophy is not to say what these norms ought to be.
- (2) Epistemology (or philosophy of science) consists of an *a posteriori* reflection upon the development and content of the sciences as they existed and have existed.

These two maxims have major heuristic consequences for the philosopher of science:

- (1) S(he) will consider that it is a major concern to pay attention to the emergence of new scientific knowledge, because s(he) believes that science is able to modify traditional philosophical images of the world and of knowledge.
- (2) Correlatively, s(he) will have no trouble in carrying out substantial historical work.

Most European, and especially continental professional philosophers of science of the 20th century fit into this description. Philosophers interested in biology are merely a special case. In France for instance, Gaston Bachelard (philosophy of physics and mathematics), and Georges Canguilhem (philosophy of medicine) are paradigmatic examples of this attitude. Since these two philosophers were the tutors, protectors and exemplary models of the most influential French philosophers of science for approximately 30 years, the predominance of “historical epistemology” in philosophical approaches to biology should not be a surprise. American philosophers of biology will generally consider such colleagues as “historians” rather than “philosophers”. But this is precisely the issue: historical epistemology claims that conceptual history of science is a major task for philosophers. In Germany, historical epistemology has deep roots in Hegel’s philosophy. In France, it is a modern term for an attitude that was clearly stated by Auguste Comte. In his lectures on “positive philosophy” (1830), Comte not only admitted that history of science was a philosophical task, but that it was the main philosophical task. Many students of my generation remember Georges Canguilhem recalling this Comtian dictum to his students.

Let me summarize this attitude in a single word. I call it “dualist” with respect to the relation between science and philosophy: science and philosophy are different, and because they are so different, it is crucial for the philosopher to understand science as it existed, exists, and develops.

By contrast, I characterize the “philosophy of biology” attitude as “unitarian”. Its basic postulate is that philosophical knowledge and scientific knowledge are not different in nature. “Philosophy” is more conceptual, because it is primarily constructed as a “metadiscourse”, but this is not an exclusive privilege. Scientists also happen to be “conceptual” and “critical”. Correlatively, “Science” is more empirical, but this is not a privilege either. Philosophers can and may confront their solutions to conceptual puzzles with the data of experience. As noted by David Hull, the absence of a clearcut distinction between “philosophy” and “science” is illustrated by the close collaboration of a significant number of philosophers and biologists in the past 20 years. For example: Sober and Lewontin, Sober and Feselstein, Sober and D.S. Wilson, Burian and Alberch-Campbell-Goodman-Kauffmann-Lande-Maynard Smith-Raup-Wolpert, Griesemer and Wade, Amundson and Lauder, etc.⁹ The same David Hull has emphasized that philosophers of biology have *de facto* demonstrated that philosophy has a role to play in biology. Of course this role is most often to clarify the meaning of scientific concepts, but this task is sometimes essential to scientists themselves.

If we now compare the typical portrait of a “philosopher of biology” with that of an “epistemologist of the life sciences”, we observe two striking differences:

A “philosopher of biology” will pay less attention to the emergence of new biological knowledge (except in the field where he helps to clarify concepts).

A “philosopher of biology” will pay less attention to history. Ideally, s(he) will avoid using historical data in order to provide an historical reconstruction. That is, s(he) will use historical data only for the purpose of testing a philosophical interpretation.

However, in practice, some philosophers of biology, perhaps many, are *also* historians of science (e.g. Michael Ruse, David Hull, John Beatty). *Also*, but *not at the same time*. Those who explicitly claim that history of science as such should be a component of philosophy of biology remain rare (Marjorie Grene, Richard Burian).

4 Examination of the Journal *Biology and Philosophy*

I will now propose an empirical test of my general comments by examining the journal that has been the very symbol of philosophy of biology for nearly 20 years. The journal *Biology and Philosophy* (hereafter *B&P*) was founded in 1986 by Michael Ruse. Some other people, like the philosopher David Hull, or biologists like Michael Ghiselin or Ernst Mayr also played an important role. I will focus on *B&P* because it has been a major tool for the historical individualization of philosophy of biology in the international community.

I have gone through the entire series from 1986 to 2002. I do not claim to have read carefully each article. I have tried to categorize the 403 articles that have been published in this period (reviews of books and short obituaries have not been taken into account). The categories that I have used do not appear as such in the journal. I have built them partly *a priori*, on the basis of questions that I wanted to test, partly *a posteriori*, that is in the course of the enquiry.

Table 1 answers two questions that were defined before beginning the enquiry. I wanted to evaluate:

- (1) The proportion of papers that were “historical” rather than “analytical”, either because the question was overtly historical (for instance an article on the contributions of German scientists to the evolutionary biology under the Nazi regime), or, more frequently, because the author tried to answer a philosophical question on the basis of an historical enquiry.
- (2) The number of analytical papers that dealt with particular biological theories (e.g. evolutionary theory), and those that raised philosophical questions of general interest about biology in general or the living world in general (e.g. papers on the notion of function, or on moral biology).

This first table was the easiest to make. Ambiguous cases were rather rare. The third column of Table 1 shows that 15% of the papers are “historical” in one way

Table 1 Distribution of papers in *Biology and Philosophy* (1986–2002)

	I Studies on particular bio- logical theories, concepts or meth- ods (analytical)	II Philosophical ques- tions of general inter- est regarding biology and the living world (analytical)	III Articles with a clear historical dimension (what- ever the subject)	Total number of papers per year
1986	5	11	3 (15%)	19
1987	22	3	2 (07%)	28
1988	10	20	2 (9%)	22
1989	23	8	2 (05%)	33
1990	8	10	3 (14%)	21
1991	7	10	5 (22%)	22
1992	13	7	2 (09%)	22
1993	8	8	4 (20%)	20
1994	3	8	8 (42%)	19
1995	10	9	1 (05%)	20
1996	9	13	2 (08%)	24
1997	9	9	0 (00%)	18
1998	11	4	8 (34%)	23
1999	14	5	6 (24%)	25
2000	10	14	4 (14%)	28
2001	13	7	6 (23%)	26
2002	17	6	3 (11,5%)	26
Total	192	152	59 (15%)	403

or another. Most of these papers are monographs on particular scientists. Very few belong to the genre of historical epistemology: very few articles make use of historical data in order to explore a philosophical problem about biology. When such papers appear, in most cases they are written by authors outside the American academic network. Since philosophy of biology as a sub-discipline originated and grew up in America, this criterion says something strong about a local style in philosophy of science. Note also the stability of the proportion of historical papers over time. Some years, however, there is a significantly bigger number of these historical papers. This is always due to the publication of a special issue in honor of a given personality (for instance Ernst Mayr or David Hull).

Table 2 provides a more detailed view of the first category in Table 1 (papers on particular biological concepts, theories or disciplines). This table also tells us something significant. Half of the papers dealing with particular biological topics are on evolution. If one adds the numbers of papers on taxonomy and on species, 72% of the papers deal with evolution *sensu lato*. If papers on ecology are added, more than 80% of the papers on particular biological theories are concerned with natural history. In contrast, the quasi-total absence of physiology, biochemistry, and biophysics is striking. The only cases of what Ernst Mayr called “functional biology” are genetics and development. Furthermore, no significant change can be seen in the distribution of this category of papers over time. The only, minor, exception is development. There have been more papers on this subject in the past 4 years

Table 2 Thematic distribution of papers in category I of Table 1 (studies on particular biological theories, concepts or methods. After 1999, it might be justified to create a new entry for “development”)

	Evolution	Taxonomy	Species	Ecology	Genetics	Other ^a
1986	3	0	0	1	0	1
1987	8	0	13	0	0	1
1988	4	0	6	0	0	0
1989	17	0	4	1	0	1
1990	3	0	3	0	2	0
1991	4	0	1	1	0	1
1992	7	1	1	1	2	1
1993	1	0	0	2	1	4
1994	0	0	0	1	0	2
1995	3	2	4	0	0	1
1996	6	0	2	0	0	1
1997	5	1	1	0	1	1
1998	8	1	0	1	1	0
1999	8	0	0	1	1	4
2000	7	0	0	2	0	1
2001	3	0	1	7	1	3
2002	9	0	2	0	1	4
Total	96 (50%)	5 (2,6%)	38 (19,7%)	18 (9,3%)	10 (5,2%)	26 (13,5%)

^a Development, psychology, physiology, cognitive sciences...

Source: **Biology and Philosophy** (1986–2002).

than ever before. The stability of data in this table suggests that the intellectual community behind the journal *B&P* is not particularly sensitive to scientific novelty in biology by and large. This seems the price to pay for an intensive cooperation with scientists on particular subjects. I have made a similar study on the major journals in history of biology. The thematic distribution of papers in these journals is significantly wider with respect to scientific domains. It is also obviously much more sensitive to scientific novelty. Thus philosophers of biology seem to be more conservative than historians of biology.

Table 3 shows the distribution of papers that deal with philosophical questions of general interest related to biology in one way or another. This classification was more difficult, because some papers could fit into several categories. Nevertheless, the overall image is striking. Nearly 60% of the papers in this category deal with evolutionary epistemology and evolutionary ethics. The proportion of papers related to evolution is in fact larger: the papers on genes and culture, teleology and mind are often “evolutionary”. Over time, this emphasis on evolution is stable. New themes have emerged, however. In most cases, they deal with cognitive science, or topics on the borderline between philosophy of biology and philosophy of mind. The net impression given by Table 3 is similar to that given by Table 2: just as philosophy of biology has been relatively sensitive to scientific novelty and diversity, it has been relatively independent of social, cultural and political change.

Table 3 Thematic distribution of papers in category II of Table 1 (Philosophical questions of general interest regarding biology and the living world)

	Evolutionary epistemology and related topics	Ethics and biology (esp. evolutionary ethics)	Nature/ culture (e.g. genes and culture)	Function, teleology, design	Reflexions on Biology in general (e.g. laws, autonomy of biology)	Other ^a
1986	2	7	1	1	0	0
1987	19	1	0	0	0	0
1988	1	3	2	0	1	0
1989	17	0	4	1	0	1
1990	3	2	1	1	1	1
1991	0	3	2	0	0	5
1992	0	3	0	1	1	2
1993	2	3	0	0	4	0
1994	3	1	1	1	2	0
1995	1	3	1	2	1	1
1996	1	2	0	1	3	6
1997	1	5	0	1	1	1
1998	2	0	0	1	1	0
1999	2	0	0	0	2	1
2000	3	3	1	2	1	1
2001	0	2	0	2	0	1
2002	1	0	0	5	1	2
Total	58 (34,3%)	38 (22,5%)	13 (7,7%)	19 (11,2%)	19 (11,2%)	22 (13%)

^a E.g.: Philosophy of mind, emotions, religion, pictorial representation in biology...

Source: *Biology and Philosophy* (1986–2002).

5 Conclusion

The study of the journal *Biology and Philosophy* confirms the claims made earlier in this paper about the regionalist turn and the historical turn. Firstly, the very existence and success of a journal like *B&P* is an obvious illustration of the regionalist turn in philosophy of science. Secondly, if we take *B&P* as fair echo of modern “philosophy of biology”, it is clear that the regionalist turn has not entailed a historical turn in the case of biology. The opposite is the case.

Before leaving the subject, I must raise the question whether a similar study could be made of the other style or tradition that I have evoked, under the name of “epistemology of the life sciences”. There are two international journals that are, in a sense, competitors to *Biology and Philosophy: Studies in History and Philosophy of the Life Sciences* (edited in Naples), and *Ludus Vitalis-Journal of Philosophy of Life Sciences* (edited in Mexico). The first periodical is mainly historical, though a significant number of the historical articles are written by philosophers. The second periodical is more obviously philosophical, but quite different in style from *Biology and Philosophy*. In *Ludus Vitalis*, the spectrum of biological topics is wider; most often, the method is the historico-critical method. For each of these two journals, it would be more difficult to make a

classification of articles than it is for *B&P*, simply because these journals are less homogeneous. Furthermore, these journals do not generate the impression of a well-organized academic community, with standard topics, and shared disciplinary norms. The style of these two journals is the typical old “HPS” style (History-and-Philosophy-of-Science), where the question whether a question is “historical” or “philosophical” one is not a relevant one, and is not intelligible for the majority of people.

Will the *B&P* style survive as a legitimate way of practicing philosophy of the life sciences? This is an open question, which can easily be related to the general state of science studies. Over the past 20 or 30 years, science studies have massively been professionalized and, correlatively, specialized.¹⁰ Philosophy of science, history of science, and social studies of science, have more or less separated. In each case, the new specialists say that they have struggled in order to “emancipate” their discipline. Philosophers of science have emancipated themselves from historians of science. Historians of science have emancipated themselves from philosophers of science. Sociologists or anthropologists of science say that they have emancipated themselves from historians of science and philosophers of science. The three branches of science studies seem to agree that the old alliance between history and philosophy of science is dead.

I have recalled these obvious points because they help to understand what is at stake in the so-called stylistic difference between “philosophy of biology” and “epistemology of the life sciences”. At first sight, it is tempting to interpret this stylistic difference as a geographical difference. That is, grossly speaking: American style versus European style. There is some truth in this interpretation, but it is not the whole story. More than geography, time is at stake: “new” versus “old”. Philosophy of biology is new in the sense of “a new discipline”. Epistemology of the life sciences is old in the sense of something that was not a discipline, with clear-cut sociological and cognitive limits. Although I have great admiration for the fertility of modern philosophy of biology, as well as for “hard” social history of biology, I doubt that increasing professionalization and specialization is the best way to guarantee the future of science studies. I tend to think that science studies are irreducibly interdisciplinary. This is their main value, in terms of knowledge, education and social communication.

Although I do believe that philosophy of biology, in the modern sense that has been analyzed here, has brought a lot of rigor in the philosophical reflection about the life sciences, I also think that it will hardly survive if it separates itself totally from other branches of science studies.

Acknowledgment I am grateful to Pierre-Olivier Méthot for his useful comments on the use of the expression “philosophie de la biologie” in France.

Endnotes

1 Whewell, 1834. On this episode, see Yeo, 1993, pp. 110–111.

2 Whewell, 1840.

- 3 Whewell, 1837.
- 4 The *Course* was published in six successive volumes from 1830 to 1842 (Comte, 1830–1842). The first forty-five lessons, including those on “biology” (40–44), were published between 1830 and 1837. The rest of the book was on “sociology”. Comte created the word “sociology”, but he did not create the word “biology”, although he was unquestionably the author who popularized it in Europe in the 1830s.
- 5 For instance Johnston, 1914; Smith, 1962. In French: Gley, 1900. See also Canguilhem observations in Canguilhem, 1952, p. 119 and Canguilhem, 1955, p. 1.
- 6 Hull, 1969.
- 7 Famous counter-examples are: Michael Ruse, Kenneth Schaffner, Alexander Rosenberg.
- 8 The distinction between the two senses of “epistemology” has been clearly analyzed by Hugo Dingler, 1936. English translation by Peter McLaughlin in Dingler, 1988.
- 9 Hull, 2000.
- 10 *Ibid.*

Bibliography

- Canguilhem, Georges (1952), *La connaissance de la vie*, Paris, Vrin, 2003.
- Canguilhem Georges (1955), *La formation du concept de réflexe*, Paris, Vrin, 1977.
- Comte Auguste (1830–1842), *Cours de philosophie positive*, 6 vols., Paris, Rouen frères (Bachelier); *The Positive Philosophy of August Comte*, trans. and condensation H. Martineau, 1853.
- Dingler Hugo (1936), “Methodik statt erkenntnistheorie und Wissenschaftslehre”, *Kantstudien*, 41, pp. 346–379; “Method instead of epistemology and philosophy science”, trans. P. McLaughlin, *Science in Context*, 2, pp. 369–408, 1988.
- Hull David (1969), “What Philosophy of Biology is Not”, *Journal of the History of Biology*, 2, pp. 241–268.
- Hull David (2000), “The Professionalization of Science Studies: Cutting Some Slack”, *Biology and Philosophy*, 15, pp. 61–91.
- Johnston James (1914), *The Philosophy of Biology*, Cambridge, Cambridge University Press, 1914.
- Smith Vincent Edward (1962), *Philosophy of Biology*, New York, Saint John’s University Press, 1962.
- Whewell William (1834), “Mrs Sommerville on the Connexion of the Sciences”, *Quarterly Review*, 51, pp. 54–68.
- Whewell William (1837), *History of the Inductive Sciences*, 3 vol., London, John W. Parker, 1837.
- Whewell William (1840), *Philosophy of the Inductive Sciences*, 2 vol., London, John W. Parker, 1840.
- Yeo Richard (1993), *Defining Science: William Whewell, Natural Knowledge and Public Debate in Early Victorian Britain*, Cambridge, Cambridge University Press, 1993.

Philosophy and Contemporary Biological Research

Claude Debru

Contemporary biological (and biomedical) research has strongly influenced existing philosophical debates and has created new ones in the recent years in epistemology, general philosophy, and ethics. My purpose in this paper is not to review these debates, ranging from reductionism to human dignity, in general terms, but to argue about a possible and closer interaction between philosophy as such and biological research as such. This kind of interaction takes place more and more today in laboratories of cognitive science or in hospitals. Some years ago, at least in France, this happened more seldomly. I wish to present some historical and perhaps also more personal comments on this practice of “epistemology in the field”, or “philosophy within science”.

After the Second World War, in the 1950s and 1960s, philosophy in France was strongly influenced by German thinking, mainly by Martin Heidegger and by Edmund Husserl. Husserl’s transcendental phenomenology became very popular thanks to the efforts of philosophers like Paul Ricœur or Suzanne Bachelard, who translated some of Husserl’s most important works, *Ideas I* and *Formal logic and Transcendental logic*.¹ In 1965 and 1966, Jacques Derrida gave lectures on Husserl’s *Lectures on a Phenomenology of Internal Time Consciousness* and *On the Origin of Geometry* at the École normale supérieure.² Husserl’s motto, “die Sache selbst”, became a basic attitude among young and eclectic philosophers, who could accommodate it with other popular philosophical schools or fashionable tendencies like Marxism (Louis Althusser), psychoanalysis (Jacques Lacan), and social anthropology (Claude Lévi-Strauss). The students who were not especially attracted by structuralist thinking as exemplified by Althusser, Lacan and Lévi-Strauss could fill the phenomenological program with epistemological and scientific content. The affinity between phenomenological and epistemological ways of thinking was felt by some philosophers in France (including Suzanne Bachelard in her book *The Con-*

C. Debru (✉)
École normale supérieure
45 rue d’Ulm
75005 Paris
France
E-mail: claudedebru@ens.fr

*sciousness of Rationality*³), and may be the reason why epistemological thinking could become a way for young philosophers to escape the internal difficulties of the phenomenological enterprise while keeping the phenomenological spirit of directing attention to “the thing itself”. Epistemology at that time was mainly represented by Georges Canguilhem, who was then the Director of the Institut d’histoire des sciences et des techniques of the University of Paris, and Suzanne Bachelard, the daughter of the late Gaston Bachelard, whose successor in the chair of philosophy of science at the Sorbonne and as the Director of the Institut d’histoire des sciences et des techniques was Georges Canguilhem himself. Later on, Suzanne Bachelard succeeded Georges Canguilhem as the head of the same Institute. The basic philosophical attitude of the Bachelards and of Canguilhem was to maintain, somehow in the spirit of Auguste Comte’s positivism, an intimate relationship between the history of science and the philosophy of science.

As a striking consequence of this attitude, science in history or science in progress was considered as a “normative” process in the sense that scientific norms could change with scientific progress without putting science in danger of losing its philosophical character of a normative process – this is the reason why a purely relativistic view of knowledge is only seldomly considered in epistemology in France, and never considered seriously in phenomenology. As a matter of fact, the idea of science as a normative process had also been investigated by Edmund Husserl before the Second World War. How could this normative process be exemplified in the history of science? The criticism (one could say the deconstruction) of fundamental ideas, which may happen before the actual discovery of new realms of reality, may induce new theoretical ideals and epistemological norms, new ideas regarding what the structure of a theory should be. It is well known that much before the appearance of relativity and quantum mechanics, Ernst Mach in his *Mechanics* had formulated a view of theoretical physics which stresses the mutual dependency of all masses in the universe (Mach’s Principle), and aims at producing by the same theoretical principle accelerated as well as inertial movements.⁴ This new program of physical theory replaced provisionally the Newtonian view of physics based on the absolute space and time frames which were considered by Mach as devoid of any empirical proof. It is well known that Mach’s philosophical ideas about physical theory as based on the concept of mutual dependency had some influence on Einstein. In another scientific domain, physiology, which Mach himself had also practiced with much success, the same observations may be made. Claude Bernard’s idea of the constancy of internal environment was a theoretical concept endowed with a unifying power in physiology including its evolutionary dimension, but based on scarce empirical evidence and devoid of any mathematical treatment. A mathematical treatment was only available at the beginning of the 20th century when the Harvard physiologist Lawrence Henderson developed the equations of blood equilibrium. This, again, was a case of mutual dependency between the components of a system, in the particular case blood considered as a physicochemical system. Epistemological ideals, or ideas of what a scientific theory should be, may thus be formulated in a similar way in different scientific domains, and before actual theoretical and mathematical treatment may occur. As norms, they may be replaced by

more realistic views, as is the case for the constancy of the internal environment, which was later considered as a rigid dogma physiologists should get rid of, and was replaced by oscillating systems and internal clocks whose workings are still not entirely understood.

As a philosopher, Georges Canguilhem was perhaps even more interested in life itself than in its science, so that he was able to take a critical look at the history of biology as a science from a philosophical point of view, which consisted of comparing the scientific conceptions of life with the phenomenon of life defined in philosophical, rather intuitive (although based on a most rigorous argumentation), and almost ontological terms. Normativity as a power of life, life defined as an unconscious position of value, were basic concepts whose speculative power was elaborated and tested in Canguilhem's famous book, *The Normal and the Pathological*.⁵ In a way, Canguilhem, the philosopher of normativity, invited his students to go back to the most basic observations. It is perhaps not a matter of chance if his teaching was consistent with the basic teaching of another philosopher of normativity, although from the very different brand of German idealism, Edmund Husserl. Some of Canguilhem's students studied medicine or biology, in order to know better "die Sache selbst", to get more or less directly acquainted, not only with science, but also and perhaps mainly with its objects. The move from abstract thinking to actual practice was rather general among young philosophers in the early 1970s in France, including the Marxist school of Louis Althusser, whose most devoted members sometimes decided to work in factories. Fellow philosophers went to factories, some went to the Medical School, I went to the Faculty of Sciences, then to the lab. This was my own way to remain at least partly faithful to Canguilhem's legacy. I learnt science not only at university classes, some of which were truly remarkable, but also while living with scientists and while doing a little research work in the lab (which I did for 18 months). Living in the lab was a radical cure against different kinds of philosophical diseases, like abstract idealism and a purely conceptual view of science, or relativistic tendencies. It was also a radical cure against the feeling of intellectual discomfort and even psychical distress which philosophy may create if pursued only as a dialectical exercise. I am fortunate enough to have maintained during many years close relationships with biochemists like Jeffries Wyman, John Edsall and René Wurmser, hematologists like Jean Bernard and Marcel Bessis, and to maintain such intellectual relationships with neurophysiologists like Michel Jouvét and Pierre Buser. From my own studies on and acquaintance with biochemistry and molecular biology, neurophysiology, or hematology, I would like to choose some topics of more general relevance to illustrate the way philosophy and science as such may interact with each other.

In the 1960s and 1970s, neuroscience developed as an interdisciplinary field of research. At some places like MIT, interdisciplinarity became, not only an ideological leitmotiv, but really a research program for neuroscience. The idea was, that physiologists of the nervous system should collaborate, not only with biochemists and anatomists, but also with mathematicians and physicists, computer scientists, psychologists, linguists, and even philosophers and epistemologists. The workings of the nervous system could not be properly understood by pure physiologists work-

ing only on the transmission of the nervous impulse. Different levels of reality, microscopic or macroscopic, should be considered at the same time. In France, similar views were held in such advanced places of neurophysiological research as Antoine Rémond's Laboratory of electroencephalography at the Salpêtrière Hospital in Paris, Jacques Paillard's Institute of neurophysiology and psychophysiology in Marseille, and Michel Jouvet's Department of experimental medicine at the Medical School in Lyon. Michel Jouvet, a physiologist and a neurosurgeon by training, is the discoverer of the so called "paradoxical phase" of sleep, which corresponds to dreaming. This concept includes a more complete description and a different interpretation for the so-called "rapid eye movement sleep" previously discovered by Eugene Aserinsky and Nathaniel Kleitman in 1953.

1 Causality and Function: The Case of Paradoxical Sleep

I was fortunate enough to enter Michel Jouvet's laboratory as a philosopher in 1981, endowed with some background in biochemistry and molecular biology, but fully ignorant of neuroscience. Later on, it became quite common to have philosophers in cognitive science laboratories. At that time, in the early 1980s, sleep and dream physiology, the subject developed by Michel Jouvet on a grand scale and in a fully interdisciplinary fashion,⁶ was facing serious difficulties due to real contradictions between empirical data on the role of monoamine neurotransmitters and especially of the serotonin molecule in sleep mechanisms. The so-called "monoaminergic theory of sleep" was based on anatomical as well as on physiological evidence. When the histofluorescence method was discovered, in Sweden, in the mid-1960s, it was realized that the cell bodies of all monoamine containing neurons in the brain were located in nuclei of the brain stem whose role in sleep was already established by lesion experiments. Physiological evidence came from pharmacology: drugs interfering with serotonin biosynthesis suppressed sleep, which was restored by serotonin precursors. Later on, in the 1970s, negative evidence began to accumulate: sleep was restored even in chronic administration of drugs, where no serotonin was present in the brain. The electrical activity of serotonergic nuclei turned out to be rather heterogeneous before and during paradoxical sleep. It was discovered also that the serotonergic nuclei could contain other neurotransmitters.

Michel Jouvet faced two major problems: how to establish causal relationships in complex systems; how to go from causality to function. He was able to make progress on both problems only in a slow way, after the rapid advances which he had made in the late 1950s and the 1960s. When I entered the laboratory under such circumstances, I learnt much, I listened much, and I did not speak too much (I just gave regular seminars on the philosophy of science for researchers, on topics like psychophysical parallelism or causality in complex systems). After 2 years, I decided to move on to real work, and I was given a subject which had already been dealt with by a technician in the lab, with whom I collaborated for 1-year and a half (we were able to confirm her preliminary results, to enlarge them and to publish them).

I tried to assimilate the physiological way of thinking. I had frequent conversations with Michel Jouvet, who showed me his numerous attempts at modeling possible solutions to the difficulties (he is always full of new ideas and was trying new experiments, the sign of a real genius). A subtle and intimate kind of interaction between the internal viewpoint of the researcher engaged on his own track and the external, philosophical and critical viewpoint was slowly established. In 1982, a crucial experiment was made, leading to the disproof of a purely serotonergic theory of sleep (although positive evidence remained and still remains).⁷ Exchange of cerebrospinal fluid (CSF) between an instrumentally sleep-deprived animal and a pharmacologically sleep-deprived animal showed that sleep could be restored in the pharmacologically deprived animal by CSF from the instrumentally deprived animal with no trace of serotonin or serotonin precursor in it. Researchers began to look for other hypnogenic substances and structures. More recently, in the late 1980s and early 1990s, important progress was made in the understanding of sleep mechanisms, neurotransmitters, and neuronal networks, thanks to the work of a student of Michel Jouvet, Pierre-Hervé Luppi and his team, leading to a much more complex picture.

Looking back at the serotonin story (an extremely interesting story from an epistemological viewpoint), the critical point was a logical and methodological one. This was fully realized by Michel Jouvet when he started his crucial experiments, showing that the serotonin molecule was only a sufficient and not a necessary condition for sleep. The inherited way of thinking in experimental physiology, which was formulated by Claude Bernard, was based on several assumptions. The first assumption was that destructive experiments reveal true causes which may be considered as necessary conditions: according to this interpretation, the rule “*sublata causa, tollitur effectus*” would mean that the removal of a putative cause leading to the disappearance of the effect is the sign that the cause is a real, proximate one. The proof is really complete when the effect is restored thanks to the reintroduction of the cause. Claude Bernard was not very good at formal logic, and perhaps he was too confident in the demonstrative power of his method. He tried to develop physiology as a science aiming at establishing the true determinism of natural phenomena. He could do it mainly by developing the surgical or pharmacological procedures of experimental physiology which interfere with the natural course of phenomena. The method of localized lesions became one of the major tools and defined the basic paradigm of experimental physiology after Bernard, but this method did not lead always to conclusive evidence. In a way, Michel Jouvet was caught in this experimental paradigm which was based on the idea of “*synthesis after fractionation*”, and was devised to divide and single out individual substances or structures in order to define their role. The serotonin story shows that this method can give only partial results. How to establish by the experimental method causal relationships in a complex system remains a true difficulty. As a matter of fact, epistemological reflection helped the physiologist to become aware of the theoretical reasons for the intrinsic limitations of his procedures. Brain research forces the physiologist to modify his methodological and theoretical views of his subject, and to create new epistemological norms.

In the particular case of sleep physiology, the difficulty of establishing causal relationships in complex systems was overcome only by multiplying the kinds of evidence. The picture is even more complicated due to the fact that the workings of the nervous system are not only excitatory, but also inhibitory, in the sense that an excitatory process can produce an inhibition in the affected cell, and that neurotransmitters can thus play an inhibitory role. How to establish, to picture, this complex dynamics of excitatory and inhibitory processes in the case of sleep and wakefulness, while taking into account the duality of sleep states emphasized by Michel Jouvet? Inhibitory molecules, like GABA and glutamate, play an important role. Other molecules, hypocretins 1 and 2, located in the hypothalamus, play also a role in the overall regulation of sleep and wakefulness. This role was demonstrated by converging genetical and physiological evidence. Genetical evidence played a major role. It was gathered from the study of narcolepsy, a genetically-determined disease, in dogs. Basic pathophysiological mechanisms of this disease in dogs were recently established by Emmanuel Mignot at Stanford.⁸ However, the relationships between these peptides, hypocretins 1 and 2, and other sleep and wakefulness regulatory molecules, remain to be more precisely defined. Still other molecules are also involved in the more and more complex picture of sleep and wakefulness mechanisms. How to model, and possibly simulate the interplay of these molecules is a task for the next generation of researchers.

In Bernard's experimental scheme, causality and function are related concepts, in the sense that removing a cause should produce serious disturbances in the function and thus help to determine it. This is true for most physiological functions but remains very doubtful in the case of sleep, due to the fact that sleep deprivation experiments did not reveal severe disturbances of the kind which could be observed for other physiological functions. The physiological function (or functions) performed by such a unique physiological phenomenon as paradoxical sleep and dreaming remains an unsolved problem, and a topic for many different hypotheses (170, according to a recent calculation of Michel Jouvet). As a matter of fact, trying to discover a function in the case of sleep lead the physiologist to abandon (at least provisionally) the purely experimental way of thinking and to think more in terms of possible models for brain functions. In the case of sleep and dreaming, as in other cases of brain physiology, trying to devise functional models for the brain was part of the interdisciplinary game of neuroscience, which involved borrowing of concepts from linguistics and computer science or from genetics and molecular biology. One of the first hypotheses regarding the function of paradoxical (or rem) sleep was proposed by a computer scientist, Edmund Dewan, in 1967.⁹ His idea was that paradoxical sleep could help programming or reprogramming genetic properties in the nervous system. Genetics, computer science and brain science began to merge at that time. Michel Jouvet had already made research on sleep ontogeny and phylogeny, which could provide arguments regarding function. In a far-seeing intuition, he defined paradoxical sleep as "genotypic arousal",¹⁰ meaning by that expression that paradoxical sleep resembles wakefulness because of its low voltage fast electrical activity, but also that during paradoxical sleep the brain is closed to any external stimulus and is opened only to its own internal functioning, whose behavioral and

genetical dimension (“oneiric behavior”) was discovered in a famous experiment. Indeed, Michel Jouvet performed lesions of the nucleus locus coeruleus of the brain stem, which turned out to have an inhibitory effect on the behavioral expression of the content of dreaming. Operated cats could thus show typical patterns of innate behavior during paradoxical sleep.¹¹

The behavioral dimension of paradoxical sleep lead Michel Jouvet to speculate that paradoxical sleep could subserve the function of reprogramming innate, genetically determined behaviors. However, suppressing paradoxical sleep did not affect basic behaviors like maternal behavior in rats. The fundamental idea of a programming or reprogramming function was kept, but the content of the hypothesis was modified due to other results on paradoxical sleep deprivation in genetically different inbred mice endowed with different performances in learning tasks. Paradoxical sleep deprivation resulted in the decrease of phenotypic variance between the mice. Thus, a possible function for paradoxical sleep could be the reprogramming of phenotypic variance, of small genetically determined differences between individuals, as proposed by Michel Jouvet. In humans, paradoxical sleep could be the guard of our individuality and even could play a regulatory, arbitration role in our continuous individuation process. I proposed this philosophical term of “individuation” in reviewing and discussing the different functional hypothesis which had been proposed by Michel Jouvet and by several other authors.¹² This problem is far from being solved, and the numerous speculations which have been made on it allow to foresee in the future an unconventional microphysiology which will be able to reformulate them, to propose new ones, and to put all hypotheses to experimental test. In such a situation, philosophy may help to formulate new ideas and perhaps even new models. This practice of philosophy is surely not philosophy “of” science, a discipline whose relationship to science as really practiced is in some cases not obvious, but philosophy in the most general meaning of a speculative activity, aiming at enlarging the field of the possible and the thinkable, and at understanding the relationships between the possible and the real, particularly in the scientific field, where the interplay between both ideas of the possible and of the real plays such an important role.

It will perhaps appear as surprising that, dealing with such a subject as sleep and dreaming, I do not venture into discussing ideas on the “mind-body relationship”, or in “philosophy of mind”, or to put it in more classical terms on psychophysical parallelism, on the cerebral correlates of conscious experience, as more and more revealed by modern imagery techniques. The field is still very immature in the case of dreaming, perhaps because psychology (cognitive psychology) is less advanced than physiology, so that correlations can be established only at a very macroscopic level which does not allow precise interpretations, and because it remains very difficult to do cognitive experiments or to discuss cognitive dimensions of physiology on dreamers, in spite of the fact that modern imagery techniques may help to overcome this obstacle. However, deep insights on the general problem of consciousness could be gained if the difference between regular dreamers and lucid dreamers (who are aware of the fact they dream) could be observed by the modern imagery techniques or more classical techniques of electroencephalographic investigation in

patients awaiting neurosurgery. I would be glad to participate as an observer in such an important scientific achievement.

2 Medical Classifications: The case of Leukemias

Going back to much more classical problems in philosophy “of” science (a phrase which I was never able to understand fully), I would like to present some comments on classification, on the particular example of leukemia classification in pathology. Which arguments could be gained from modern classification techniques in hematological pathology regarding the “nominalistic” versus the “realistic” conception of classification? When Jean Bernard and Marcel Bessis asked me, as a philosopher, to consider the problems of leukemia classification in order to participate in the improvement of classification methods, I was struck by the variety of classification methods and criteria in use, by the pragmatic content of classification regarding the standardization of diagnostic and treatment, and above all by the fascinating problem posed by “unclassifiable leukemias”, a problem dealt with by Jean Bernard and Marcel Bessis in the first issue of their journal *Blood Cells* in 1975.¹³ The primary concern of these hematologists was to devise a classification scheme which would allow them to find regular pathophysiological mechanisms and general laws underlying the differentiation of leukemias. Their hope was to produce a natural classification which could play in pathology the role played by Mendeleïev’s periodic classification of chemical elements in chemistry and physics. The chemical classification reveals an order expressing structural laws valid at the atomic level. However, these laws were discovered by atomic physics, with the chemical classification playing only a kind of confirmation role. As to hematology, the extremely well organized community of hematologists entertained (and still entertains) intense discussions about classes and classification, due to the introduction of new techniques (a major move occurred when morphological and biochemical data were supplemented by immunological data), and new treatments (which provide data which are included in classificatory schemes). However, pathophysiological mechanisms (which are still known in a very imperfect way, and in a better way only for a small number of leukemias, like promyelocytic leukemia) are studied by genetics and molecular biology, with classification playing only the role of a preliminary basis.

Leukemia was defined by Rudolf Virchow, who coined the term in 1847, as a disease of the developmental process of blood cells. Already at the beginning, the qualitative meaning of the term, and its extension (the cases it could designate) were much discussed. As a matter of fact, the term is used to designate diseases which are not always diseases of blood cell differentiation, characterized by a proliferation rather than by a differentiation process. For example, chronic lymphocytic leukemia is an accumulative rather than a proliferative disease. The question of the unity or multiplicity of leukemias, which was asked from the beginning, remains thus an open question, even for apparently well-defined types of leukemias. The terminol-

ogy remains a unitary one, due to a long history. The facts are increasingly differentiated, thanks to new biological techniques.

In the following, I would like to argue in favor of a more realistic interpretation of classes in pathology, and perhaps also in favor of a reformulation of inherited philosophical dilemmas in terms which would be more relevant to contemporary scientific practice. This more realistic interpretation may be unexpected, if one considers the rapidly changing character of classifications in pathology. However, it may be supported by biology itself. The point is first, that pathology affects a differentiation process by which nature itself creates well-defined differences up to a final state, starting from an undifferentiated state. And second, that pathology affects this differentiation process in a way which keeps its most important characters alive and valid. This double paradox was well perceived by Jean Bernard, who noticed in 1975 that there is something quite illogical in founding the classification of pathological states on the classification of normal states, by giving to pathological cells the same names as to normal ones (mainly: myeloid leukemias and lymphoid leukemias, a difference based on the main lines of normal differentiation of white blood cells, or myeloblasts and lymphoblasts, which are normal precursors of mature cells); second, that this illogical way of classifying leukemias had quite happy consequences regarding prognosis and treatment of leukemias for many years.¹⁴ More recently however, with new biological criteria and new treatments, the practice of subdividing existing classes increased, leading to a more and more realistic picture of leukemias.

The case for a nominalistic conception of classification could be supported by the rapid changes which were observed in some cases, like hairy-cell leukemia. In the 1950s, hairy-cell leukemia (HCL - a purely descriptive, morphological term) was considered as a variant of chronic lymphocytic leukemia (CLL). In the 1960s, phase-contrast microscopy provided results which lead to a different view of the disease as a specific disease, an infiltration of histiocytes through the bone-marrow, due to the resemblance of hairy cells to histiocytes. Later on, immunology and cytogenetics showed that these pathological cells belonged to the family of lymphocytes, an unexpected return to the earlier concept of a chronic lymphocytic leukemia. The introduction of a new treatment, interferon, lead to the different conclusion, that HCL and CLL are not the same disease, since they respond differently to the treatment. A more recent treatment, 2 - chlorodeoxyadenosine, reunifies HCL and CLL. This picture of a rapidly fluctuating classification of a disease is quite spectacular. However, it is the sign of insufficient knowledge, and the sign of the purely empirical basis of treatments. The final word was given when pathologists realized that the early morphological picture was sound enough.

Another, more serious argument in favor of a purely nominalistic, thus provisional and finally pragmatic view of classification is the evolving character of pathology itself, due to the well-known phenomenon of progression of the disease, which consists of the fact that the same case of a disease evolves from one type to another. Laszlo Lajtha gave a striking example of that progression: a case of acute lymphoblastic leukemia, once treated, changed into acute myeloblastic leukemia, again treated. Finally, the patient died from erythroleukemia (leukemia of red blood

cells). As a comment, Lajtha quoted Lewis Carroll: Why give them names, if they don't answer them?¹⁵ Medical knowledge makes progress thanks to unexpected cases revealed by new techniques (immunological techniques, in the following example). In the 1980s, cases of biphenotypic leukemias (leukemias endowed with both myeloid and lymphoid characters) were increasingly described thanks to the use of monoclonal antibodies.¹⁶ How to interpret these cases? By which processes could these mixed types arise? It was felt that these cases were diseases affecting the early stages of differentiation. They were described as "lineage infidelity" or "lineage promiscuity". In spite of their quite peculiar character, these cases were not discussed in terms of classification. They were discussed in terms of what they could bring to the understanding of the early stages of differentiation.

Progress of medical knowledge thanks to the use of new techniques lead to further differentiations. An interesting case is provided by chronic lymphocytic leukemias (CLL). They are subdivided according to the types of lymphocytes, B-lymphocytes or T-lymphocytes, which are affected. However, two different circumstances modify this beautiful and simple picture. First, in B-CLL, which amounts to 95 % of all cases, T-lymphocytes function is also affected. Second, B-CLL must be further subdivided. B-CLL turns out to be a rather heterogeneous group with no real coherence between the different criteria of biological and genetical characters and clinical outcome (survival). The picture is increasingly complex, and in such a situation classification is urgently needed for successful treatment. However, the conclusion which can be drawn from the present picture at the epistemological and medical levels is not that pessimistic. Federico Caligaris-Cappio wrote recently:

CLL is a heterogeneous disease, with some patients having a long survival and never requiring treatment and others running an aggressive course and demanding intensive therapy. Until recently, it was not possible at diagnosis to assign patients to either group. The observation that two subsets of patients may be recognized on the basis of the presence or absence of somatic mutations of immunoglobulin genes has changed the rules of the game. The presence of somatic mutations has been correlated with CLL clinical course and response to therapy [...] These data, together with those provided by cytogenetical studies, indicate that dissecting the clinical heterogeneity of CLL is another feasible goal. Such a distinction would allow an individually tailored and presumably more successful treatment.¹⁷

This judgment by a well-known specialist indicates clearly that the goal of classification remains the treatment of the individual. It makes it possible to ask a slightly different question: should classification reach the level of the individual at the molecular level? Does variation in pathology reach the level of the individual? There are many reasons to think so. Unclassifiable leukemias could be understood in this way.

There remains truth in leukemia classifications, which makes them not a purely nominalistic and variable exercise. Like the great divisions in the tree of life, which remain extremely solid, the great divisions in the classification of leukemias keep their solidity from the fact that the differentiation process is itself well defined and is more and more precisely understood at the molecular level, even if the borders may be occasionally crossed at early stages. Pathophysiological mechanisms interfering with the normal differentiation process include genetical, chromosomal abnormalities. They include also more and more a view of the plurality of causes, internal and

external (as was proved by the statistical analysis of Burkitt's lymphoma, which shows the involvement of a virus, a gene, and malaria). A purely nominalistic theory of classification remains unable to account for the reality of the differentiation process, which nature performs constantly, up to the level of individuals, and to account for the pathophysiological mechanisms which govern the division in types and subtypes of the diseases in a way which is more and more specifically understood.

Regarding the need, felt by prominent hematologists like Jean Bernard and Marcel Bessis, to improve classification methods, to make them both natural and rational, and not purely empirical (as is still the case with the use of treatments in classification schemes), the answer provided by an epistemological and historical analysis of the subject would be that classification does not reveal by itself etiology or pathophysiology but is part of the progress of etiological and pathophysiological research. As such, classification is only one part of the scientific game. This means also that inherited philosophical questions like nominalism versus realism of classes are much in danger of losing their scientific relevance. Clearly, there has been important debates regarding methods of classification throughout the history of biology. They seem to be less vivid nowadays at a pragmatic level. Jean Bernard's and Marcel Bessis' questions to the philosophers were entirely justified at the scientific and medical level (no successful treatment without well-established classes). The progress of medical knowledge brings the hope that classes may reach the level of individuals, and that individually tailored treatments based on a thorough biological, molecular knowledge of each case may be available, so that classification, the creation of large groups, would become only a preliminary step.

Finally, I would like to promote the idea of philosophy as an interpretative, reflective and speculative activity which can be practiced within science as well as about science. I tried to show that this is possible (the same could be said about technology, particularly about biotechnology). It would be a great satisfaction to see young philosophers who would be ready to engage in this long and fascinating journey.

Endnotes

- 1 Edmund Husserl, *Idées directrices pour une phénoménologie*, French translation, introduction and notes by Paul Ricœur, Paris, Gallimard, 1950; Edmund Husserl, *Logique formelle et logique transcendentale. Essai d'une critique de la raison logique*, French translation by Suzanne Bachelard, Paris, Presses Universitaires de France, 1957.
- 2 Edmund Husserl, *L'origine de la géométrie*, French translation and introduction par Jacques Derrida, Paris, Presses Universitaires de France, 1962.
- 3 Bachelard, 1958.
- 4 Mach, 1883.
- 5 Canguilhem, 1943.
- 6 Jouvét, 1992.
- 7 Sallanon et al. 1982, p. 145.
- 8 Fromherz and Mignot, 2004, p. 479.
- 9 Dewan, 1965, p. 365; 1967–1968, p. 296.
- 10 Jouvét, 1972, p. 270.
- 11 Sastre and Jouvét (1979), p. 982.

- 12 Debru, 2006.
- 13 Bernard, 1975, "Introduction", p. 3.
- 14 Ibid.
- 15 Lajtha, 1975, p. 54. Cf. Debru, 1998, p. 278.
- 16 Parentesis et al. 1983, p. 63.
- 17 Caligaris-Cappio et al. 1993, p. 603.

Bibliography

- Bachelard Suzanne (1958), *La conscience de rationalité. Etude phénoménologique sur la physique mathématique*, Paris, Presses Universitaires de France.
- Bernard Jean (1975), "Introduction", *Unclassifiable Leukemias, Blood Cells*, 1, Marcel Bessis and Georges Brecher (eds.), Berlin, New York, Springe-Verlag.
- Caligaris-Cappio F. et al. (1993), "The Nature of the B-lymphocytic Leucemia", *Human Leukemias: Re-examination of Dogmas. Blood Cells*, 19, pp. 601–613.
- Canguilhem Georges (1943), *Le Normal et le Pathologique*, Paris, Presses Universitaires de France, 1966; *The Normal and the Pathological*, trans. Carolyn R. Fawcett in collaboration with Robert S. Cohen, New York, Zone Books, 1989.
- Debru Claude (1998), *Philosophie de l'inconnu. Le vivant et la recherche*, Paris, Presses Universitaires de France.
- Debru Claude (2006), *Neurophilosophie du rêve*, 2nd ed., Preface Michel Jouvét, Paris, Hermann.
- Dewan Edmund M. (1965), "The P (programming) Hypothesis for REM Sleep", *Psychophysiology*, 4, p.365.
- Dewan Edmund (1967–1968), "The Programming (P) Hypothesis for REM Sleep", in E. Hartmann (ed.), *Sleep and Dreaming*, International Psychiatry Clinics, vol. 7 no. 2, Boston, Little Brown, pp. 295–307, 1970.
- Fromherz S. and Mignot E. (2004), "Narcolepsy Research: Past, Present and Future Perspectives", *Archives italiennes de biologie*, 142, pp. 479–486.
- Husserl Edmund (1913), *Ideen zu einer reinen Phänomenologie und phänomenologischen Philosophie*, Halles, Niemeyer; *Ideas: General Introduction to Pure Phenomenology*, trans. W.R. Boyce Gibson, London, Allen and Unwin, 1931.
- Husserl Edmund (1929), *Formale und transzendente Logik*, Halle, Niemeyer.
- Husserl Edmund (1939) "Die Frage nach dem Ursprung der Geometrie als intentionalhistorisches Problem", *Revue internationale de philosophie*, 1, 1939, pp. 203–225. *L'Origine de la géométrie*, French translation and Introduction by Jacques Derrida, Paris, Presses Universitaires de France, 1962.
- Jouvét Michel (1972), "The Role of Monoamine-containing Neurone in the Regulation of the Sleep-waking Cycle". *Ergebnisse der Physiologie*, 64, pp. 166–307.
- Jouvét Michel (1992), *Le sommeil et le rêve*, Paris, Odile Jacob; *The Paradox of Sleep: the Story of Dreaming*, trans. Laurence Garey, Cambridge, Massachusetts, MIT Press, 1999.
- Lajtha Laszlo (1975), "Answers to Submitted Questions", *Unclassifiable Leukemias, Blood Cells* 1.
- Mach Ernst (1883), *Die Mechanik in ihrer Entwicklung historisch-kritisch dargestellt*, Leipzig, F.A. Brockhaus; *The Science of Mechanics. A Critical and Historical Exposition of its Principles*, trans. Ph.E.B. Jourdain, La Salle, Ill., Open Court, 1960.
- Parentesis J. et al. (1983), "Biphenotypic Leukemia: Immunologic and Morphologic Evidence for a Common Lymphoid-myeloid Progenitor in Humans", *The Journal of Pediatrics*, 102, p. 63.
- Sallanon M., Buda C. and Jouvét M. (1982), "Restoration of Paradoxical Sleep by Cerebrospinal Fluid Transfer to PCPA Pretreated Insomniac Cats", *Brain Research*, 251, p. 145.
- Sastre Jean-Pierre and Jouvét M. (1979), "Le comportement onirique du chat", *Physiology and Behavior*, 22, pp. 979–989.

Part V
Philosophy of the Behavioral and
Cognitive Sciences

What is a Mental Function?

Joëlle Proust

To Marie-Claude Lorne

It is a remarkable and puzzling fact that, for over a century, psychological and biological research have been exploring the development and functional characterization of brain/mind activity in almost totally separate and non-interactive ways. It cannot be denied, however, that the human brain is the result of evolution in the brains of other mammals. The subcortical structures of mammals present anatomical, neurochemical and functional homologies, and these suggest largely similar mechanisms for emotion, perception and action. One of the reasons for the lack of concern by experimental psychologists with biological issues may derive from a premature, and thus largely sterile, nature-nurture controversy. While it was difficult in the recent past to understand how genes might interact with the environment in expressing themselves, the notion of epigenetic development is now better understood in its precise mechanisms, although much remains to be discovered.¹ The nature-nurture controversy, however, has overshadowed other important points. Even if it is recognized that human behavior largely results from individual and socio-historical interaction with the environment, such interactions can hardly be understood if the general constraints that the species confronts, given its bodily structure, general needs, and physical-social environment, fail to be grasped.

In order to approach the notion of a psychological (or mental) function in a principled manner, we need to understand, in general terms, what a mental function is, and how it relates to brain evolution. To achieve this, we will first need to summarize how teleological discourse has been “naturalized”, i.e. how functional explanation has been defined in purely causal terms (Sect. 1). We will also need to characterize how mental functions differ from other organic functions, and consider the causal constraints that are exerted over evolutionary time on this type of function (Sects. 2–4). In order to discuss this issue, however, we will first have to examine the respective roles of genes and development in regulating adult cognition, and determine the cor-

J. Proust (✉)
Institut Jean Nicod
CNRS UMR 8129
29 rue d’Ulm
75005 Paris, France
E-mail: j.proust@ehess.fr

rect methodology to use in addressing the question of mental function. Various definitions of mental function correspond to different methodological viewpoints, that we will discuss successively: Evolutionary Psychology² and its modular approach to function, Evo-devo and the idea of a developmental view of function, and neurocognitive theories, with their notion that neural growth determines function. The goal of this paper is to define the concept of a “mental function” that meets the general constraints that apply to the concept of a biological function. Its ultimate aims are to understand how psychology as a theoretical field is articulated with biology, and identify which methodological requirements are entailed by this articulation.

1 The Concept of Function

We say that a structure of type X in species S has function F , when the following three conditions are fulfilled:

- (1) Structures of type X in species S typically produce F in inclusive context C .
- (2) Structures of Type X are inheritable or reproducible.
- (3) Structures of type X exist because they have adaptive value (because those organisms that have produced F thanks to X in the corresponding inclusive context have been/are more likely to reproduce than those that have not).

There are three different causal relations involved in our definition of “function”. The first is the idea of a *causal disposition*, that is a mechanism that tends (or used to tend) to produce a given consequence. For example, a muscle tends to contract, a property that has many possible usages, from pumping in or out, to locomotion and prehension. The second is a *reproductive mechanism*, that allows the device in question to be “copied” or emulated either in the same organism (for further use) or in its offspring. The causal mechanism here does not need to be specified: it ranges from genetic reproduction to epigenetically stabilized features, (as with muscles), learning and cultural scaffolding (as with mindreading or planning). The third causal relation is the crux of the dynamic-causal explanation of function, through the notion of heightened fitness: the device in question has been selected, reproduced and/or is presently used *because it gives the organism a better chance for survival*, through the precise effect it produces in the relevant context.

This causal condition is sometimes seen as a “historical” form of causality; but actually, things are more complicated. It can be interpreted either in retrospective, historical terms, as the “etiological view” of function recommends, or in dispositional, time-neutral terms, as the propensionist view suggests.³ We will not discuss here the merits of these interpretations, which correspond to the respective needs of evolutionary biologists (interested in the evolutionary history of traits) and of anatomo-physiologists (interested in the causal role of a functional element) as well as to what Mayr called ‘functional biology’.⁴ Suffice it to say that this third causal link is a form of causation by consequences immersed in a recurrent selective process. For our present discussion, conditions 1 and 2 need to be spelled out in more detail.

1.1 *Inclusive Context*

The notion of an inclusive context, introduced in condition 1, is a very important factor that modulates what is called “the categorical basis” of the corresponding disposition. Context is not only the background in which a function operates, it plays a structuring causal role in the functional effect. It is simple to see why. A physical (or biophysical) disposition owes its causal capacity to there being nomic relations between properties engaged by the disposition and the environment in which it is exerted. For a substance to be soluble in water, water is not needed: as a pure possibility, the disposition is relational and context-independent. For this disposition to be exercised however, the substance must actually be plunged into water. If the contexts of a functional device are sufficiently modified, the corresponding disposition fails to be convertible in a causal process.

Moreover, given that the structure, in virtue of condition 2, needs to be reproducible in new organisms to count as a functional item, context needs to be stable at least in its causally crucial dimensions, to allow the structure to reliably produce its effects. Context interacts with function in two different ways, i.e. according to two types of constraints (by “constraints”, are meant the contextual conditions that have to be present for a given functional disposition to be exercised). There are two kinds of such causal constraints that a device must satisfy to produce a functional effect *F*. These constraints have to do respectively with the external or internal context in which the device operates.

External context constraints are the nomic biophysical relations that allow the energy flow between the device and the relevant part of the environment. Let us consider two examples. (1) muscles are built over the course of development in given conditions of gravity. Actually gravity causally contributes to the development and operation of muscles (Thelen and Smith, 1994). (2) Social learning occurs in given conditions of appropriate social, communicational and emotional environment, etc. The kind of feedback that the system receives about the type of social relations, and the value of these relations for it, is causally crucial for learning to occur.

Internal context constraints are the relations between the functional device producing *F* and the other functional subsystems of the organism with which the structure producing *F* interacts in order to produce *F*. For example, a muscle works in relation with other functional elements: the skeleton dictates the shape and location of the tendons and of the skeletal muscles, which are themselves involved in moving various body segments according to the various demands of motor tasks. Learning how to read minds presupposes a hierarchy of other mental capacities. To read a mind, you need (*inter alia*) to have conceptual-inferential ability, a linguistic ability including syntax and a semantics for mental states, and an evolved ability to teach and to learn. For any function, there are upstream – or precursor – functional elements that need to be present for the function to be assembled or exercised (downstream functions will also in turn depend on it to develop, but they do not constitute a constraint for their precursor). Bill Wimsatt has theorized this functional asymmetry in terms of a relation of *generative entrenchment* between a functional effect

and the preconditions that make its development and normal operation possible.⁵ The solidarity at any given time between the various elements contributing to the *F* effect is what allows us to speak of a “system”, such as the musculo-skeletal system, or the mindreading system.

1.2 *Reproducibility*

The second important aspect of our definition has to do with the concept of reproducibility or inheritability. For evolution by natural selection to occur, the variants that, in a given population and environment, increase their bearers’ fitness must be heritable traits. The concept of function, being part of a fitness-based selective reproductive process, involves the intervention of a reproductive device. Quite obviously, one important process that underlies this transmission is the genetic mechanisms of reproduction. Transmitting human hearts (normally) presupposes sexual reproduction. There are however also non-genetic mechanisms that contribute to the recurrence of a functional trait. Some are environmental – such as those ecological parameters that control development, motivation and growth in similar ways in various individuals. Some, and probably most, are cultural – such as the vast number of learning practices that permanently “reproduce” specific skills and associated tools in new brains – and hands; linguistic communication that creates new tokens of a thought pattern in new individuals; bodily communication that creates new expressive means in the recipients; various prosthetic technologies allowing artifacts “resembling” the dysfunctional structure that is to be replaced or otherwise repaired and/or enhanced in order to maintain the organism’s life. These various kinds of examples are meant to underscore that no identity or exact duplication is required by condition 2.⁶ Every case of transmission has to allow for contingent developmental accidents as well as for interfering stable dynamic patterns that may modify the outcome. We will see later, however, that such variance is a welcome feature that can be exploited and monitored in a dynamic system.

1.3 *The Causal-Teleological Condition*

We will not comment in detail on condition 3, which has been discussed at length in heated debates among evolutionary biologists. Stephen Jay Gould and Richard Lewontin have ridiculed the “Panglossian” tendency of adaptationists to identify each single organic feature as a functional trait, much like the Leibnizian Pangloss of Voltaire’s *Candide*. Their objections to an oversimplified application of the concept of function are well taken and are devastating for the naive teleologist that is dormant in us. In order to apply condition 3, as they are right to insist, it must be shown that no other explanation is at hand than the causation-by-consequence explanation (other explanations include developmental or architectural constraints

that cannot fail to apply – and thus do not need to be “selected” –, or ‘non-selective’ genetic explanations, like genetic drift). A methodological recommendation follows about applying condition 3 in a parsimonious manner: no function should be attributed without documentation of a relevant selective history. This kind of consideration motivates the next question we want to raise.

2 Mental Function: The Modular View

Having provided a general definition of what a function is, we now need to find the specific difference that will allow us to focus on mental functions. What are the features that make an organic function a mental one? In other terms, what is the disposition (or set of dispositions) that can be copied or reproduced such that having it would allow a system to develop capacities of a mental type (condition 1)? What are the general constraints – internal or external – that apply to them? What are the specific “copying” or “reproductive” mechanisms that implement them (condition 2)? And how do the consequences of the disposition about fitness explain that the mental disposition is adaptive, i.e. has been selected (condition 3)? In sum, are we in a position to identify mental functions?

Although the concept of a function is of common use in psychology and in cognitive science, experimental psychologists have rarely addressed these three questions in any systematic way. It is often observed that experimental psychologists studying a capacity in mature adults rarely attempt to understand the development and full scope of the capacity they are studying. They are aiming to characterize behavioral regularities, not to question the way they were established, which functions they have outside the lab, and which neuronal systems realize them. Although there are, as we shall see below, deep reasons for “functional agnosticism”, there are also more shallow ones. The functional terms used in experimental psychology, such as “perception”, “working memory”, “motivation”, etc. are used to characterize, in broad functional terms, systems of informational processing mechanisms that are logically required, given basic assumptions on what a mind should be. The term *function* is used to refer to a causal role in a set of informational processes relating input and output. Although this usage is *prima facie* pragmatically sufficient to conduct experiments and theorize about the mind, it cannot provide a foundation for psychology as a science; far more is required than a broad causal role “to cut nature at its joints”: the problem with these causal attributions is that both the stimuli and the responses are classified in terms of commonsensical, rather than causally relevant, “natural kinds”; they actually fail to be robustly involved in the functional roles that they are supposed to have (for example, it is debatable that there is such a thing as “visual perception” in any strict sense of the term, given the multimodal organization of the perceptual system in mammals).

It may be worth emphasizing that this use of the term “function” did not originate in biology, but rather in rational taxonomy, which from Aristotle on has decomposed the mind into general, wide purpose ‘faculties’: perception has the purpose

of extracting information (perceiving), memory helps store and retrieve it (learning, remembering), motivation allows wanting, acting is the faculty of converting knowledge into goal-directed behaviors, and emoting allows one to create social bonds and to communicate. The various dispositions cited as constituting a mind fail to qualify as biological functions, however, as long as they are not justified by the type of reasoning we have sketched above: no internal or external constraints are investigated concerning how the disposition is supposed to deliver its functional effects, no homologies are invoked across related species to explain how corresponding traits are inherited. Finally, no explanation is given of why the device or mechanism has been selected and reproduced.

It is in reaction to this kind of broad causal-role (non-biological) view of mental function that Evolutionary Psychology has developed. The idea is to systematically identify the teleological condition (condition 3 above) that explains why an information-processing disposition is present. In virtue of condition 3, a given psychological function exists if the specific informational pick-up and the computational transformation between input/output that it effects have been selected because they solve a specific adaptive problem. Where traditional views on the mind identify general-purpose, content-free mechanisms (“learning”, “reasoning”, “emoting”, etc.), Evolutionary Psychologists claim that a specialized content – that which constitutes a specific “essential adaptation” – must organize the very design of an informational mechanism. This in turn suggests that a cognitive architecture has to be composed of many modules, each one having been selected to solve a specific adaptive problem.

As Jerry Fodor defined the term (Fodor, 1983), a module is an informational processing device that automatically, quickly and effortlessly, transforms inputs of a domain into readily usable outputs. Modules however are “informationally encapsulated” in the sense that they use only the information available in their own domain of specialization (for example: visual or linguistic processing); they cannot modulate their outputs by using the various types of information stored in other modules. A common example of this informational encapsulation is offered by such robust perceptual phenomena as the Müller-Lyer illusion: this illusion persists even when the perceiver is aware that the two segments are in fact equal. Whereas Jerry Fodor took modules to be “peripheral” entities, delivering their outputs to a non-modular, central processing unit, Evolutionary Psychologists maintain that functional specialization extends right up into reasoning and other higher-level processes. This fully modular view of cognition, according to them, is the only way to solve the “frame problem” that is, to bypass the computational explosion that would paralyze a general-purpose system confronted with real-world complexity. Indeed a module is the evolutionary solution to combinatorial explosion *as a general selection pressure*.

There are several ways of solving the frame problem: reducing the dimensionality of the problem situation can be achieved by using either domain specific algorithms, or non-standard information-processing models. An alternative way – as we shall see below – is to show that the problem never actually arises for biological systems: it is only generated by an erroneous view of how these systems relate to their “informational” environment. On this alternative view, the brain imposes structure on its environment in a way that reduces its dimensionality. For

Evolutionary Psychologists, the problem does arise, and dedicated computational modules are the multiple, but automatic, responses that evolution found to solve it. “Darwinian algorithms” or “cognitive programs”⁷ are automatically activated each time a certain recurrent adaptive problem is encountered. Given that there are many such problems, the mind encompasses many different modules, from cheater detection and mate selection, to food choice, habitat choice and theory of mind.

A notorious difficulty with the notion of an evolutionary module is that of justifying the “adaptive high-resolution maps” of the various mechanisms that collectively constitute the mind. Three types of methods can be used to construct such maps. One is the “reverse-engineering” approach: starting from the problems that our human ancestors had to deal with, find which precise cognitive programs would optimally respond to them. The second consists in testing the hypotheses formulated in this a priori way. The crux of the method consists in creating alternative experimental conditions, in which the distal “adaptive” dimensions of the task are made to compete with more proximal computational explanations of the observed performances. For example, if there is such a thing as a “cheater detection module” whose activity is elicited by a “social contract” situation, experiments will be designed to contrast performances when a detection of a rule violation is presented in a social contract garb, and in an abstract reasoning garb (as in the original Wason task). If one can experimentally show that performance is greatly facilitated in the “adaptive” presentation, and cannot be better explained by some other adaptation or by a general purpose capacity, then the case can be made for a task-specific module being active. A third kind of evidence makes use of neuropsychology and psychopathology: if double dissociations between, say, theory of mind, on the one hand, and cheater-detection on the other, can be observed in brain-lesioned subjects or patients with autism (one kind of performance being maintained while the other is disturbed), then the hypothesis for there being (at least) two different specialized modules for social reasoning is comforted.

Let us pause to consider what Evolutionary Psychology has to tell us about mental functions. First, the idea is that there are as many mental functions as there are specialized modules. Only some of them have been discovered, mostly those that correspond to the Fodorian peripheral modules. Evolutionary Psychology however suggests that all the modules that constitute our reasoning and decisional capacities use computational capacities in a contextual and parsimonious way. Therefore, old-fashioned psychology and neuroscience study only apparent functions: for the responses they describe do not fulfill the three conditions above. (1) They fail to identify the contextual, motivated or adaptive character of information processing. For example, people do not use a “mental logic” to detect rule violation; they use, rather, a faster and more frugal algorithm that bypasses *modus tollens*. (2) They don’t explain how this disposition is acquired and reproduced (it certainly is acquired neither through the teaching of logic, nor through innate logic). (3) They don’t explain how the disposition is there.

To capture the level at which it is relevant to speak of a mental function, Evolutionary Psychologists rely on David Marr’s (1982) three-level distinction of a research agenda in cognitive science. The “computational theory” has to determine

what the goal of the computation is. The level of the “algorithm” has to determine how the input, its transformations, and the output are represented. Finally, the hardware level has to determine how representations are implemented in the brain. By adopting this trichotomy, they choose to look at “mental function” from a functionalist viewpoint (this is not, appearance notwithstanding, a tautology: “functionalism” refers to a view of the mind according to which mental states are definable in terms of their causal-representational relations, independently of their cerebral realization). Mental function is taken to belong to the computational level: function has to do with the goal of a computation: the goal must be clear before we can look for the cognitive processes that subserve it. The “hardware” level, finally, has at most the role of providing additional, optional evidence for the existence of a function. This view of mental function has been very popular among philosophers and psychologists, and more generally among functionalists: if only the causal relations, captured in computational terms, between inputs and outputs are relevant in psychology, then a neuroscientific analysis of the particular way in which these computations are implemented is only of marginal interest.

Many objections have been leveled against Evolutionary Psychology, often inspired by a strong anti-biological view of human psychology. We need only consider here those objections that share the view that Evolution is relevant to understanding mental function. There are three main objections of this sort.⁸

2.1 *Methodology*

The first is methodological. Reverse engineering seems to bring back the worst forms of pan-adaptationism. Although some information processing mechanisms must have been selected for their effects, it is difficult to prove that the teleological condition 3 does apply in all the cases where an adaptation seems to be present. Some might be rather exaptations,⁹ or architectural consequences that have no direct teleological explanation. A currently controversial example is the computational apparatus underlying human language, in particular syntax. Hauser et al. (2002) suggested that syntax might have evolved for reasons other than language (like number use, navigation, social cognition). If such were the case, syntax would be an exaptation rather than an adaptation, in spite of the obviously useful effects of a language faculty in our species. Pinker and Jackendoff (2005) on the other hand, argue that the complexity of the interconnected mechanisms that syntax involves have the earmarks of adaptation, which suggests that syntax is an adaptation for communication. Another example is the possible functionality of postural synchronous sway among cooperative speakers:¹⁰ although this sway might be adaptive to facilitate communication among participants, it might also be simply the exaptive effect of intrinsic rhythmic patterns in speech production.

It is furthermore not obvious that evolution of mental function would always proceed by finding and using *the* optimal design for solving an adaptive problem. How efficient a cognitive mechanism is, depends on various dimensions of the adaptive problem: centrality of impact (how well does it solve the problem given the mental architecture in its

previous design), tolerance for mishaps, temporal constraints, and proportion of cognitive resources that it recruits. Given the necessary trade-offs between these variables, it is to be expected that the hypothetical selected modules should often be sub-optimal in accuracy, while offering acceptable solutions given the time and effort saved. This objective fact however works in favor of overincluding behaviors among adaptations, which is a reason to exert caution in using this line of reasoning.

2.2 *Co-Evolution*

Some adaptive changes in one species (for example: color pattern in birds) are such that they do not generate a major selective pressure on other species, nor directly affect the physical environment. This however does not seem to be the case for evolutionary changes that are driven by social environments, as is the case for human cognition. According to the Machiavellian Intelligence Hypothesis, any increment in social predictive capacity is bound to have repercussions within and also beyond the species.¹¹ In the primate group, social pressures have led to adaptations for representing and predicting representational and predictive capacities in others (which allowed a “theory of mind” to emerge). The technological and cultural capacities that humans have developed as a social species on the basis of their cognitive capacities have also been constantly transforming the physical environments in which they construct their niches, to such an extent that all the other biota and associated ecosystems have been more or less directly affected. These transformations to the environment constitute new pressures directed at more specific adaptations.

This has led many theorists to conclude, as do Sterelny and Griffiths (1999)¹² and Proust (2006), that there is no invariant environment to which the lineage is adapted. Change in group size, change in population structure, and change in available resources strongly modulate the adaptiveness of altruistic or Machiavellian dispositions in a way that may differ from individual to individual, according to the strategy implemented in the group at large. Technologies of various kinds, as well as human cultural practices, also strongly affect the physical environment in which they have to survive and reproduce. If this is true, cognitive adaptations should be taken to accommodate an evolving rather than a stable environment. Mental function should accordingly not be taken to correspond to one recurrent adaptive and “essential” problem (cashed out in terms of stable social and physical configurations). It should rather be characterized as a set of capacities allowing the organism to cope with changing, largely unpredictable environments.

2.3 *Predictability and Non-Flexibility*

The switch from the evolution of mechanisms to the evolution of evolvability of mechanisms constitutes an a priori reason to doubt that modules might be an efficient response to social-cognitive pressures: algorithms specialized in dealing with

others in such practical matters as resource sharing or mate choice would be rigid responses to a changing world, not flexible answers to ever changing social constraints. They would also be particularly vulnerable to exploitation by others, given that predictions are made easier when the information relied upon is modular and specialized. The only way to prevent rigidity and predictability would be to have many possible modular responses for any given situation. But in this case, some form of central control of the modular responses should be expected; we would lose automaticity and speed, and we would be confronted with the computational explosion that threatened the general learning account.

In sum: Evolutionary Psychology has interestingly tried to respond to the requirements that an informational device must fulfill to qualify as a mental function. (1) It explicitly aims to identify the relevant context in which a given informational process was selected. (2) It takes genetic transmission to explain how “Darwinian algorithms” are reproduced. (3) It accounts for the presence of these specialized informational processes by a historical-causal (i.e. “etiological”) application of teleological condition 3. Central objections however blame reverse engineering methodology; they pinpoint that the appropriate selective context might be characterized in dynamic rather than in factual/discrete terms. Furthermore, they observe that there is presently no clear evidence that modules are genetically transmitted. Finally, the selection of shallow, fast and automatic modular processing seems to be incompatible with Machiavellian constraints, which are part of the inclusive context in which functions are stabilized.

3 Mental Function: The Developmental View and “Evo-Devo” Biology

Several developmental approaches to evolution, in particular Evo-devo and Developmental System Theory (DST), have generated important ideas that are crucial to our discussion of mental function(s).¹³ Evo-Devo is the field of evolutionary biology that examines how development has evolved under evolutionary pressures, and how it retroacts on evolution. One of the central concepts of this field is the concept of modularity, which turns out to be of critical importance to understand the basis of mental functions. Developmental system theorists on the other hand question the orthodox view – that genes control the progressive functional specialization of brain areas in interaction with the physical and the social environment –, a position that, according to them, takes mental functions to be predetermined in the genetic code. For DST theorists, in contrast, genes are themselves in part controlled by a variety of epigenetic chemical processes such as the DNA methylation system. In addition, inheritance mechanisms are not exclusively genetic, but extend to a variety of chemical, cultural, behavioral and physical phenomena.

For the proponents of a developmental approach, therefore, what counts as an adaptive problem cannot be read from ancestral environment alone; it can only be known from the structure of what they call the “developmental modules” that

structure the growth of the organism under study. A developmental module is a “region of strong interaction in an interaction matrix” (Griffiths, to appear). Developmental modules do have causal effects on other modules (they allow differentiation to occur in neighboring parts); they have however a higher degree of internal interaction, which allows them to be viewed as independent building blocks in the development of an organism. The adaptiveness of such a modular development is similar to the adaptiveness attributed to mental modules: in a modular architecture, the effects of mutations are more local; thus their disruptive effect is less likely to harm the whole developmental process. Although some of these mutations may be quite specific and apparently limited, the outcome however may be extremely important, in particular because of the hierarchical and temporal nature of the “developmental cascade”. As Griffiths emphasizes, double dissociation is also a consequence of this modular developmental architecture: a developmental module can be identified on the basis of its being selectively impaired without the other modules being perturbed.

Now why should a developmental module be more relevant to mental function than the mental modules identified by Evolutionary Psychology? The response is straightforward: because developmental modules *directly map* the selection pressures. To understand how such a mapping is possible, and how the developmental modules finally connect with mental modules, it is necessary to return to the circular causal process, in virtue of which niches are causally shaped by populations while also constituting causal constraints to which populations have to adapt. When a co-evolutionary process affects both a specific population and its ecological niche, each co-transforms the other, creating ever changing adaptive problems for the organisms living in that niche. Thus, if you want to identify the selective pressures, look at the developmental modules (e.g. look at how the cells develop, how they differentiate in cascade in interaction with the environment). The mental modules, those that express the stabilized, mature organism, cannot however be directly inferred from the interaction between the species and its environment. Mental modules are much more elusive functional entities, for they result from the conjunction of two factors. First, each organism has a specific developmental pattern resulting from the specific timing of its encounter with relevant events and properties. Second, every mind will be the outcome of an optimization of its informational processes in light of cost-benefit compromises. For example, the degree of granularity that categorization reaches within a domain will depend on past affordances and individual training. There is no “general” environment with normal affordances and normal training procedures. Variation is the rule.

Griffiths and Sterelny thus blame the Evolutionary Psychologists for their lack of sensitivity to what they call the “grain” problem. Let us take an example. What is the level of specialization of a given cognitive program, such as the “cheater detection program” studied by Cosmides and Tooby. Is it designed to detect a cheater? Is it designed to be more emotionally engaged in a task when the risk of a social contract violation is present? Is it to memorize learnt rules when they are more salient? Evolutionary Psychologists claim that a mental function is a unique solution to one independent pressure; reverse engineering is able, starting from the problem, to derive the cognitive mechanisms subserving it. Such a method is indeed successfully applied

in the case of morphological traits: the stability of the relevant selective pressures and the obvious relation, for example, between bone length, gait and body mass may allow one to form precise functional claims. In the case of psychological mechanisms, however, there is no a priori indication of the correct grain size.

Only a developmental approach, however, can tell what is dissociable as a separate task; furthermore, the individual's specific interaction with the environment and the ensuing learning process are what determine his resulting psychological organization – i.e. the task – and environment-dependent connections between informational processes that prevail for that individual. For that reason Paul Griffiths (to appear) takes mental functions to consist neither of “virtual modules”, which need not coincide with development, nor, as we will see, with neuronal modules. This notion of “virtuality” is important, in that it points to the fact that mental function is associated with a kind of replication that cannot be described independently of the context of development.

3.1 The Case for Homologies: A Tension Between Two Requirements?

DST theorists as well as philosophers of biology¹⁴ have emphasized that homologies should play an essential role in determining biological functions in general, and mental functions in particular. In contrast with analogies, homologies characterize structures, which are similar inside phyla because of a shared ancestry (analogies exist between structures that are similar across phyla, without a shared ancestry). In order to justify the existence of a given function, as many biologists have argued, it is not sufficient to cite a causal link between an adaptive problem and a specialized mechanism; it is also necessary to show that the trait under consideration – the computational mechanism controlling the adaptive behavior – is the result of descent from earlier adaptations. This was, indeed, one of the main messages of Lorenz and Tinbergen's ethological view of function: a comparative approach is necessary to justify an evolutionary explanation of current behavior.

More recently, authors of different obediences, such as Griffiths (1997), Matthen (2000) or Panksepp and Panksepp (2000), have used, in arguing against Evolutionary Psychology, the role of homologies within cladistic reasoning: “functional classifications can be used to group cladistic units together in a way that allows the theorist to express generalizations about the evolutionary process that apply to many different lineages” (Griffiths 1997, p. 217). Homologies are indeed helpful for two basic reasons. First, they allow us to trace back how a structure has evolved along a phylum, from precursor to new forms. This allows us to make function attributions that are more strongly rooted in the organism's anatomy and in the fact that the trait was transmitted because of the additional fitness it conferred on its bearer. Second, they allow comparisons that are not based merely on superficial similarities in behavior; for example, if there is homologous brain circuitry that is activated for fear in macaques and in human beings, there is a *prima facie* strong reason to say that this circuit globally has the function of helping the system to detect and respond to

danger; this even allows us to make precise hypotheses about the kind of danger to which the circuitry most efficiently responds (snakes, predators etc.). When no such cerebral homologies are at hand, we are in no position to establish that two analogous behaviors have one and the same function – that is, have been selected because of the particular response they allow, and on the increased bearers' fitness that ensues.

In the case of human psychological functions, homologies have recently played a major role in functional reasoning. For example, although only human beings have linguistic competence, it has been shown that there is a homology between premotor area F5 in monkeys and Broca's area (Brodmann's areas 44 and 45) in humans (Rizzolatti and Arbib 1998). These areas contain assemblies of neurons that are activated in primates both for executing and simulating manual actions and facial gestures (a "mirror-system" for action), while Broca's area is involved in phonetic, syntactic and semantic language processing. This homology certainly does not show that monkeys can or could speak, given adequate training; but it suggests that the mirror-system might be a crucial pre-adaptation of language; the "missing link" between primate non-speaking communicators and human speakers might thus consist in an extension of the mirror system, linking it to exclusively human imitative abilities. An action proto-language might have emerged in humans, according to this reasoning, as a capacity for syntactically combining semantically interpreted manual gestures – a capacity implemented in Broca's area (Roy and Arbib 2005). Although mirror neurons were not originally "meant" (selected) to be used in communication, they might have later become a crucial piece of proto-language production and understanding. This example shows that homologies can be used to understand the evolutionary lineage of an adaptation within a phylum – rather than simply helping dissociate mental functions by distinguishing their underlying neuronal realizations.

To summarize, the requirement of finding homologies in function attribution, seems to entail that psychological functions should fail to be identifiable when no homological brain structure is. Condition 2 in our definition of function is here becoming more central than condition 3. Griffiths also maintains, however, that psychological functions may include "virtual modules", that is: dissociable patterns of performance that do not correspond uniquely to separate neuronal systems. If some or most psychological functions are such "competence patterns", which are shaped by the demands of the current developmental environment rather than by an inherited brain anatomy, it is not clear why we should retain the homology requirement in psychology. If flexibility and plasticity prevail in the mental, it might seem that psychological functions do not have to be identified with inherited anatomo-physiological structures. The theoretician's burden, in this case, is to give substance to the inheritance condition 2 on function in non-structural terms.

3.2 The Ambiguity of What is Called a Function

The preceding discussion brings again to the fore the ambiguity that plagues most uses of the expression "mental function". In current terminology, "neuronal systems" refer to anatomically specified task-specific activations of neuronal assemblies. The

very existence of these neuronal systems seems to most theorists to offer a sufficient guide to mental function. It is common parlance to discuss “the” functions of “the mirror system”, of the “dorsal stream” of perceptual processing, or the role of the amygdala in processing emotional information in perceived faces etc. But the problem here is twofold. First, it is not obvious that what is identified as a causal role (as revealed by a systematic task-specific pattern of activation in the adult brain) actually corresponds to a mental function as defined by condition 3. What is reflected in brain imagery may not be carving out ‘a’ function: it may in fact express either one or several functions (as called for by the particular task), one or several “exaptations”, or simply stabilized interactions between brain, culture and physical environment that do not have a function in the defined sense (without being *inherited* capacities – being there *because* they have this particular effect).

Conversely, what should lead us to expect that bona fide mental functions necessarily correspond to the activity of dissociable, separate neuronal systems? If Griffiths is right to think that virtual modules do not have to be associated with a fixed neuronal substrate, then mental functions “in the strict sense” might be realized by one cerebral substrate or another across subjects, according to the circumstances of their development – where developmental timing pattern, specific early training, injury, or local cultural habits, might play an essential role. Perhaps, in other words, the apparent neuronal fixity of modules is an effect of cultural/environmental homogeneity rather than a solid fact about the brain. If these reflections are on the right track, then the discussion of homology for mental function might have to be upgraded to a higher-order architectural level. We should not expect to find lineages of strictly dedicated, task-specific zones, but rather very general domains, such as emotional areas, versus spatial, instrumental, and episodic knowledge areas and the effects of their action, none of which can function independently of the others. The notion of a virtual psychological module is thus perfectly compatible with the view that the whole brain constitutes the only neuronal module there is.

4 Mental Function: The Neuronal Growth Approach

We saw above that Evolutionary Psychologists have used Marr’s trichotomy in such a way that mental function turns out to be dissociable from brain development, physiology and anatomy. Many biologists and DST theories, however, have objected that homologies have to be present for sound functional attribution: they are the biological earmarks of functional structures. Fortunately, a solution to the homology puzzle is in view. Powerful new ideas on the structural homologies subserving mental functions have emerged in the last few decades from the neurosciences, which come surprisingly close to the theoretical intuitions of the DST theorists – although no clear explicit cross-influence is noticeable in the respective literatures: the mind is built from the developing brain; the dynamics of the brain can only be understood against the background of the dynamics of the organism’s environment; neuronal growth is the key to understanding why certain forms of brain activity are selected for their effects.

A major difference in accommodating these ideas, however, concerns the respective roles of learning and of brain dynamics. For “neural constructivism”,¹⁵ *learning* is what *guides* brain growth: learning induces changes in the brain structures involved in learning. For neural selectionism (also called brain Darwinism), the reverse is true: *brain development drives* learning, by the neuronal competition and selection that it generates. Before coming back to this important difference, let us introduce and comment further shared intuitions on mental function among neuro-cognitive theories.

For both theories, the brain is a “representational device” (*representation* here being taken to mean that the brain states are correlated with other states and properties, about which they carry information). It is widely recognized that representational development in ontogeny is characterized by “U shape” patterns. Children begin by performing well on some task, then they typically undergo a period of failure, by overgeneralizing their earlier knowledge, until they finally come up with a new stable, more robust, and extensive ability.¹⁶ This phenomenon points to the fact that the brain is *nonstationary* – its statistical properties vary with time. Both theories agree that this nonstationary character deeply affects the ways we should think about mental properties, for the structures underlying acquisition themselves change over time. Later representational stages are not simply refinements of earlier stages, but involve large-scale reorganizations as a consequence of structural changes in the learning mechanisms. Both reject the functionalist interpretation of Marr’s trichotomy. Both see the brain as using feedback to regulate its own development. Finally, and crucially, both claim that the cerebral vehicle of mental function is dynamically shaped by the very process that allows mental function to emerge, *and* according to a set of mechanisms that are “meant to” – i.e. shaped by evolution in order to – let that function emerge. In other words, the functional analysis that allows us to understand mental function is referred to the neural vehicle that implements mental activity. This does not constitute a “change of level”, as the authors take it that cognition is actually developing at the cell level and across neural populations: dendrite segments and dendrite structure are carrying information; the architecture of cognition can therefore only be understood through the architecture of the developing brain. This claim is a major breakthrough in theorizing about mental function, a breakthrough that needs to be explored in all its consequences.

The traditional genetic explanations of brain function, as we saw above, assume the brain to be “genetically informed” of the general kind of environment it is going to be living in. Further gene-environment interaction in epigenesis is supposed to fine-tune adaptation. In contrast with this view, the neurocognitive theorists are claiming that the brain has the disposition to be dynamically shaped to allow survival in a changing world. The genes work by biasing the brain/world interaction early in development, and not – or not only¹⁷ – by conveying information about which rules or algorithms to apply. The maturational program coded by genes now works only or mainly as a biasing factor: it leads certain areas of the developing brain to be sensitive to particular kinds of inputs.

In both theories, although they may not express their results this way, the intrinsically dynamic (and, as I will argue later, teleological) form that results

from developmental brain/world interaction is that of a hierarchy of adaptive control loops. In other terms, the neurocognitive theories reject the distinction between a cognitive and an implementation level because the very same type of mechanisms (generate and test procedures, for the selectionist, or “constructive”, that is, environment-controlled growth, for the neural constructivist) characterize the development of cognitive learning and of the growth or selection of neuronal cells. This latter point is very important, and needs to be discussed in more depth in each theory.

4.1 Selectionism : “A Radically New View of the Function of the Brain”(Edelman, 1987)

The selectionist theories, introduced by Edelman (1987) and Changeux and Dehaene (1989), adamantly reject the idea that mental function could fail to correspond to neural organization. Changeux and Dehaene (1989) suggest, rather, reinterpreting Marr’s trichotomy so that it refers to different organization levels *within the nervous system*. The most basic architectural level they describe is the single cell level, with a functional differentiation of the axon, dendrite and synapse. At that level, the function of the neuron can already be deemed ‘cognitive’: it is to transform input into output, in virtue of specific patterns of electrical and chemical properties that carry information. A single neuron is already performing a computational task (the program level); it is following an algorithmic process, and does so according to specific physical properties (molecular properties of the synapse and of the membrane). There is therefore no “ontological” autonomy of any one task-level, as functionalists claim, but a relation of “*co-dependence*” among levels. The constraints of the synapse and the membrane determine, in part, which computations can be performed, as well as which kind of goal they can serve. Reciprocally, serving a goal modulates both the computational and the physical levels, and helps stabilize the physical properties of the cell. A second anatomical layer encompasses “circuits”, i.e. neuronal assemblies of thousands of cells organized in well-defined structures, i.e. presenting task-dependent synchronous firings. A third layer is constituted by the “metacircuits”, i.e. relations of neuronal assemblies. Finally the traditional mental faculties are taken to roughly correspond to various of these metacircuits.

Functional agnostics will be quick to ask how such an organization emerges in the first place. The response is that a recurrent two-phase process is responsible for brain organization and learning. An initial exuberant growth of neural structure, leading to an overproduction of synapses, is followed by a selective pruning back of connections. There are successive waves of this sort of growth and selection from birth to puberty, each wave presenting in succession “transient redundancy and selective stabilization” . “One has the impression, writes Changeux, that the system becomes more and more ordered as it receives ‘instructions’ from the environment”.¹⁸ This impression, in the author’s view, is justified; indeed the function

of the brain comes down to that: stabilizing those dynamic patterns that have high predictive value, while suppressing those that have low value *given the environment in which development is taking place*.

In summary: learning occurs by selection (as a consequence of brain/world interactions), but only if neurons and their synaptic connections already exist. This succession is objectively justified by the fact that learning presupposes selection. Bouts of learning can accordingly be analyzed, as the authors claim, through some version or other of Herbert Simon's "generate and test" procedure. Neural proliferation produces variety; neural pruning selects those variants that have been more often activated through feedback from the environment (by suppressing their less successful competitors). Transitions between levels of organization result from a generalized and hierarchical stabilizing effect of "generate and test" procedures with re-entrant feedback loops within larger populations of neurons.

The most striking aspect of this theory, from a philosophical point of view, is that a "causal-teleological" explanation of adaptation, goal or function, is now taken to be common to the phylogenetic evolution of the brain, to developmental (ontogenetic) evolution, to the workings of the mature brain, and implicitly (as we will show below) to the representations that the brain structures carry. "An analogous Darwinian scheme" as Changeux and Dehaene (1989) put it, is at work within the brain as it is within evolution at large: brain structures have evolved neural growth/learning processes that mimic the teleological patterns that populations of organisms are subjected to.

4.2 *Neural Constructivism*

This alternative theory takes its inspiration from Piaget's constructivism, i.e. the view that mental representations are constructed through an action-guided, ongoing internalization of environmental structures. As we saw above, careful observation of human ontogeny shows that learning in each domain is highly discontinuous. Neural constructivists explain the dynamics of learning in development by a progressive growth of dendrites according to the interaction of the brain with perceptual input (in each modality: visual, auditory, proprioceptive, etc.). The fundamental differences with selectionism involve two claims: (1) that dendrite growth (and diversity) is exclusively controlled by the environment, rather than dually by endogenous and exogenous influences;¹⁹ (2) that the immature cortex is taken to be initially equipotent (as brain damage in early development has little or no detrimental effect on mental function). Evidence from brain plasticity suggests that brain function is, as we saw above, a matter of general evolvability rather than of specialized adaptations. For neural constructivism, the actual functional organization of the mature brain depends entirely on the external constraints that the brain needs to learn: "It is the differing pattern of afferent activity, reflective of different sensory modalities, that confers area-specific properties onto the cortex – not predispositions that are somehow embedded in the recipient cortical structure".²⁰

As a consequence of claims 1 and 2, neural suppression plays only a minor role in brain development. The structuring force consists rather in the creating of neural connections under the influence of incoming data/stimuli. The mechanisms that are hypothesized to generate brain tissue growth and, more specifically, dendritic arborization, seem to involve local releases of neurotrophins, i.e. feedback signals that are delivered post-synaptically and are thus activity-dependent signals.²¹ As a consequence of these constructive, bottom-up mechanisms, the cortex is “enslaved”, that is, fully controlled, by the periphery. Mental function thus consists primarily in “enslavability”: it involves the production of flexible, adapted responses to varying environmental constraints as well as to changing body size. As in the selectionist model, the constructive model associates mental function with a hierarchical brain architecture; hierarchical representations result from cascades of environmental constructivist influences working from cells to assemblies onto circuits, thus building representations of increasing complexity.²²

Constructivists however are more insistent than selectionists in considering that mental function can only be understood as the particular (and somehow contingent) outcome of development. In their view, domain-specific competences of the Darwinian algorithm type (cheater detection, snake detection, etc.) can be seen as evolved “mental functions” only if one forgets the whole developmental process that generated them. Actually, they are the result of repeated processing of initially domain-general mechanisms, which have turned out to be more often used for specific inputs: *domain-relevant* mechanisms are thus progressively turned into *domain-specific* mechanisms, as a result of their particular developmental history (Karmiloff-Smith 1992).

In summary, we see that although the two neurocognitive theories under review are similarly focusing on the dynamics of development and its cascading effects on brain structure and function, they have symmetrical views on the relations of brain and environment. Selectionists see the brain as imposing structure, through its own innate “biasing” agenda, on an unstructured world. Neural constructivists reciprocally see the world as enslaving the brain by imposing on it spatio-temporal patterns of reactivity and sets of representations.

The variety and complexity of the brain processes engaged in learning suggest, however, that the two views might in fact have to coalesce into some encompassing “hybrid” theory: regressive and constructive mechanisms might in fact concurrently be engaged in development – even though, existing evidence can still be argued to favor one camp over the other.²³

To conclude this section, it may be helpful to summarize the preceding discussion in terms of our analysis of function (see Sect. 1). Concerning condition 1, i.e. an existing disposition (or a set of dispositions) that can be copied or reproduced such that having it would allow a system to develop capacities of a mental type: neurocognitivists answer that it is the disposition of the brain to grow in a way that is sensitive to its developmental environment (more exactly, to the computational demands that it involves). Note that this description of the brain disposition itself includes embedded functions: for example development is itself selected for, and may therefore also evolve as such, which in turn will affect the very disposition of the brain to structure itself. The strength of the neurocognitive proposal for mental

function, as compared to modular views, is that the general constraints that exert pressures on brain evolution are made fully explicit in the model. The external constraints are the environmental conditions that differentially affect neural growth; the internal constraints are represented by the set of existing neural assemblies and their interconnections that allow the brain to grow in ways that are in part predetermined by the existing circuits and metacircuits. Condition 2, which concerns the processes that are used to reproduce the disposition in other organisms, is dually constituted by genetic reproduction and by the mechanisms that allow stabilization of the environment through human intervention. Finally, condition 3 posits that the consequences of the disposition to regulate brain growth is correlated with a capacity to extract information and process it in the way that is the most flexible given the overall external and internal constraints, and has been selected because of this correlation. These theories therefore see mental function as a progressively differentiated, but initially global, capacity to store previous dynamics in existing brain matter in order to predict the environment and to adjust to it. In a currently fashionable style, one could say that the distal function of the brain is to orient its growth so as to “resonate to” the environment – or to be “dynamically coupled” with it in a flexible way. We will elaborate this view in the final section.

5 Mental Function or Functions?

The mental representations that are built as an outcome of (creative or regressive) neural growth have a set of functional properties that are necessary ingredients of a working mind. I call these properties “functional” because they are *entailed* by the distal function of the brain that was just spelled out. As a consequence of the selectionist/constructivist model, these representations must be predictive, recombinable, modifiable, robust, and have a descriptive/conative polarity. We will first examine each of these features in order to see which cognitive capacities or developmental constraints refer to them. We will then be in a position to approach a more speculative question: How does recognition of these functional dimensions affect our view concerning how many mental functions there are?

- a) *Predictability* is an outcome of the control structure of the developing brain. Representations are predictive, in the sense that the feedback used to construct a representation corresponds to a state of the world whose temporal properties are context-relativized. In other words, the brain takes advantage of the contextual cues to predict what comes next in a structured way (keeping track of perceptual cues as well as temporal sequence to reach probabilistically reliable predictions). For example, in an ordinary human environment, meals occur as a succession of routines with a certain dynamic pattern. The brain needs only store a small set of cues to know that it will soon be time for lunch.
- b) Representations can be *recombined* in ways that tend to be less and less specialized with development, in virtue of the very capacity of the mind/brain, as a dynamic system, to generalize knowledge acquired in a specific domain.²⁴

- c) *Modifiability* is also a result of the constant re-evaluation of acquired informational structures through feedback. Flexibility applies both to learning procedures (during development) and to thought contents (throughout life): both are adjusted, monitored and reorganized in a self-organizing way. Representational flexibility, however, should occur in a differential way, constrained as it is by generative entrenchment; this differential flexibility is in keeping with the fact that a representational system, such as a brain, is a nonstationary system with time-dependent properties.
- d) *Robustness* is a necessary feature of representational stability in a dynamic world.²⁵ In genetics, mutational robustness refers to an organism's phenotype remaining constant in spite of mutation. By analogy, representational robustness is a property of representations and representation systems in virtue of which they preserve a stable core in spite of being transmitted or generalized to new contexts, or in spite of contextual change. This property of robustness is of major architectural significance given, again, the role of potentially varying, unstable feedback in neural growth. Although as we have seen, learning how to learn is a crucial adaptation to changing environments, the environment in which learning occurs may also be customized to regulate robustness, and lead to a form of environmental selection comparable to niche selection in evolutionary biology.

Two sources of feedback-induced robustness come to mind. One consists in using the body and its own dynamics as a general model for other dynamic phenomena. Bodily motions can be used as a way of representing dynamically social as well as non-social events through overt or covert simulations.²⁶ Bodily gestures, such as pointing in joint attention, can be used to reduce noise in communication; they can also facilitate recall as well as thinking (considerable bodily information goes to structuring "abstract" linguistic symbols).²⁷ The other consists in using the social and physical environment as an external device to shape children's development. For development to occur, there must be, as Lev Vygostky and later Jerome Bruner insisted, a process of scaffolding through which a child is guided by parents and teachers, allowing him/her to move forward into a zone of proximal development that the child could not reach alone.²⁸ An efficient scaffolding is one that optimizes the "fit" between the informational content being transmitted (a skill or piece of knowledge) and the cognitive properties of the receiving mind. The scaffolding may include, besides the "executive control" of an adult mind, a timely and sufficiently attractive presentation of the task, as well as spatial and sensorimotor cues that allow multiple representation of the problem space. This latter "fit" is regulated in turn by another selection/creation process that applies socially to representations. This process creates what Dan Sperber calls "cultural cognitive causal chains".²⁹

Bodily postures, ways of moving as culturally shaped, as well as cognitively adapted environments are therefore also selected and maintained in dynamic coupling with representational systems (each stabilizing the other). They contribute to a considerable degree to representational robustness. This kind of co-evolution might help explain why there are cultural invariants, such as folk-biology, folk-psychology, or religion: the representations that prevail in the socially constructed mind/brain are those that are easier to acquire and transmit, and that fit more snugly the various

emotional/motivational demands of the developing brain. Reciprocally, the social environments and bodily practices that allow an easier grasp (perceptual and cognitive) of the associated representations should evolve under the renewed demands of mind/brains that have depended, for their development, on those environments.

In sum, the brain needs a culture to grow into an organized semistable structure, much as its cells need oxygen to survive. *Pace* Evolutionary Psychologists, however, the existence of cultural universals – such as folk biology or folk psychology – does not necessarily point to an innate, genetically predetermined modular organization; these can also be explained in terms of the brain's permanent activity – cognitively resonating to a social/biological/physical environment possessing globally similar constraints.

a) *Conative-descriptive* polarity is an architectural requirement that any cognitive organism has to fulfill: it must deal with information in two complementary ways – extract it, in perception, and use it, in action. In neurocognitive models, this polarity is already implemented at the cell level, each neural cell receiving input and producing a response – (graded) firing or no firing – according to the input. Motor behavior is the organism-level response through which adequate coupling with the environment is performed in a flexible and integrated way (momentary needs are integrated into a single goal-oriented behavior at a time). In human cognition, the higher level at which this polarity controls behavior is the prefrontal cortex level. This structure is a highly adaptable structure that is constantly configuring the system to address current concerns, taking into consideration various time-dependent constraints.³⁰

We are at last in a position to offer an answer to our question: how does this analysis of mental function accommodate traditional notions such as the five faculties (perception, action, memory, affect-desire, reasoning) or the more recent distinction between selective attention, perception, working, semantic, and episodic memories, planning and action, emotion etc.? *If the function of the brain is to ensure a cognitive dynamic coupling with its environment, driven by inputs and biased by innate motivations, is it still biologically justified to distinguish separate mental functions according to their causal roles?*

A conservative response would defend the view that a simple rational task analysis of what the brain needs to accomplish (i.e. develop into a flexibly learning, sensitive organ able to control behavior) shows how justified the old wisdom was. Information has to be picked up (perception), retained (memory), and it has to be used (action), in a way that is context sensitive (motivation/emotion) and able to combine inferentially with existing knowledge (reasoning). It cannot be disputed that even though these venerable divisions of the mind turn out to be theoretically ill-grounded, they might be practically valuable in allowing research to develop when no principled explanation of mental function is available – or is consensual enough to organize research.

A more speculative and daring view, one that the developmental approach and the neurocognitive models of brain growth have made plausible, is that it is far from obvious that the brain is structured into separate, anatomically distinct faculties of the kind proposed. Multi-reentrant loops dynamically connect perceptual areas with

motor areas, memory, and emotion sites. Any task has to cut across all of the “faculties”, and will foster further development of their interconnectedness; such interconnectedness *constitutes* what learning that task consists in. To know how to do X, you must do more than “simply remember” – a capacity that would involve a pure form of “procedural memory” – you must also evaluate the benefit (a subcortical achievement), you must have perceptual access to X-related objects, and your brain must have some somatic experience corresponding to how it feels to do X. Performance in action is thus constantly modulated by motivation and emotion as well as by perceptual input and memory, and actually every such contribution is closely associated with others, either by cortical or subcortical connections. Vision psychologists acknowledge that perceiving involves acting with the eyes: muscles regulate vision, and retinas perform saccades that explore the relevant parts of objects of interest. Reciprocally, action psychologists acknowledge that action is constrained by the visual or the auditory spatial properties of the effect of the action (Simon effect), not to mention the essential role of perceptual feedback in the control of action. This functional connectedness suggests that the attempt to treat the traditional faculties as distinct mental functions is not biologically justified.³¹

An additional argument for a global and dynamic notion of mental function comes from pathology. Contrary to the traditional faculties model, it is not the case that blind subjects are those whose eyes are lost, that mad people are those whose reason is perturbed, that aboulic people are those that cannot act etc. In fact, what cognitive pathology has been showing is that there is no such thing as a “mental dysfunction” that would coincide nicely with the impairment of a brain structure and its associated hypothetical role; impairments rather affect subjects in many different ways, suggesting that the traditional syndromes are only phenotypically similar but in fact causally complex.³² Far from justifying the existence of these separate but collaborative functions, pathology reveals the importance of development timing, and interaction between early acquisitions and motivation in organizing the adult competences (Karmiloff-Smith 1998). Furthermore, as recent research on Williams Syndrome and autism has shown, different subjects may use completely different strategies to solve the same problem or perform the same cognitive task. For a theory of mind task, for example, some subjects rely more on their semantic memory, some on perceptual cues, some on instrumental reasoning. These various strategies were presumably selected as a consequence of the particular pattern of plasticity that their developing brain has been settling into, given the genetic or environmental constraints it had to face (Karmiloff-Smith 2006).

If, as we want to claim, adaptive control is the general “mechanism” or rather, type of mechanism, that characterizes mental function,³³ an evolved causal mechanism that shapes brain growth as well as genetic expression, representational selection and propagation of culture, then we need to look at mental functions as either additional adaptations or exaptations to adaptive control. Kim Sterelny (Sterelny, 2003) for example has suggested that mental skills are specifically meant to cope with what Sterelny calls “informational translucency”, a property of social or physical environments in which the cues are less reliable or can be manipulated by predators. In such environments, there is a cost in mining information (because of the

risks incurred in exploring the presence or value of the cues) and/or in acting on it (when the cues are not reliable, the action becomes ineffective). In such contexts, it becomes important to devise strategies not only for reaching external goals, but also for extracting and using information. One of the most important functions of human culture is to help young organisms acquire the capacity to assess informational quality and to restore transparency whenever it is possible and useful to do so (by changing either the internal or external environments).³⁴ Sterelny explains robust multi-modal cueing and representation decoupling (allowing separate storage of alternative, incompatible representations of the same kind of situation³⁵ as responding to such pressures. Do these skills represent new cognitive functions? In Sterelny's view, these skills have rather to be analyzed as domain-general responses elaborated through culture. The idea of a "cumulatively engineered epistemic environment" refers to the view that the educational or imitative procedures that characterize our human cultures are designed to allow individuals to acquire "entrenched skills". Given that such skills are a product of learning provided with what we earlier called "scaffolding", it is arguable that they do *not* constitute additional cognitive "functions"; they are rather the expression of how the brain preserves representational robustness in uncertain or socially demanding contexts.

In Proust (2006b and in print), I suggested that another important dimension of robustness consists in predicting one's own capacity to predict. Higher level prediction defines metacognition, a control process that is endogeneously applied to the brain's predictive self-evaluative processes. This second order prediction has been studied in metamemory, a field of interest to teachers; more recently, its importance has been discovered for neuroeconomics, where predictive valuation processes are made to apply to the value of predictors of rewards. Should metacognition be taken to represent a distinct mental function? I would want to claim, rather, that the associated skills point to dimensions of adaptive-predictive control systems that are inherent in mental architecture. They are not separate functions, but constitutive aspects of this overarching, uniquely adaptive, predictive, "dynamic coupling" function of animal minds of a given complexity.

Acknowledgement This article is dedicated to my former thesis student and researcher in the philosophy of biology Marie-Claude Lorne, who read a previous version of this article and offered precious comments to me. She took her own life on september 22, 2008. I am mourning her, as do all her friends and colleagues from HPST, IJN and other institutions where she worked.

I thank Dick Carter for his linguistic revision, as well as for his critical observations. Research leading to this chapter was conducted in the context of a fellowship on Embodied Communication, at the University of Bielefeld (Germany) during the Summer 2006. I wish to express all my thanks to Günther Knoblich, Ipke Wachsmuth and Scott Jordan for interesting discussions.

Endnotes

- 1 See Oyama, 2000, Jaenisch and Bird, 2003.
- 2 The capitalization is meant to distinguish a specific school in evolutionary biology from the general field (referred to by the same, non-capitalized, term).

- 3 For a defense of the propensionist account, see Proust, 1997.
- 4 See in Mayr, 1961/1988, "Cause and Effect in Biology", pp. 24–37.
- 5 On this concept, see Wimsatt, 1986 and Griffiths, 1996.
- 6 Dawkins' meme theory claims that ideas are identically replicated from brain to brain, and undergo a quasi-darwinian selective process comparable to genes. Sperber's epidemiology of representation, however, does not postulate that a copying process is involved. Rather, there is repeated production by the recipients' own informational processes, which allows variants of the same type to be generated by the same epidemiological process. See Sperber, 2006.
- 7 Cosmides and Tooby, 1994, p. 73.
- 8 For a general critical approach of Evolutionary Psychology, see Buller, 2005 and Panksepp and Panksepp, 2000.
- 9 An exaptation is an adaptation where the effect currently performed by the corresponding trait does not coincide with the effect explaining why the trait was primarily selected. Exaptations, arguably, might still count as functions: although initially selected for effect A, they might now have effect B, which would explain the disposition of the corresponding structure to raise its bearer's fitness and thus be passed on to offspring in the future. But usefulness of an exaptation does not automatically mean that it will actually be passed on: it may also have deleterious effects given the inclusive context in which the organism has to survive.
- 10 Shockley et al. 2003.
- 11 Whiten, 1997.
- 12 Sterelny and Griffiths, 1999, pp. 328 sq.
- 13 For a clear presentation of these two endeavours, see Robert et al. 2001.
- 14 Griffiths, (in press); see also Amundson and Lauder, 1994.
- 15 Karmiloff-Smith, 1992, Thelen and Smith, 1994, Quartz and Sejnowski, 1997, Christensen and Hooker, 2000.
- 16 For precisely documented examples, see Karmiloff-Smith, 1992.
- 17 As we shall see below, some selectionists recognize that genes might make available an initial repertoire of dedicated processes, on which selection will suppress the non-relevant ones.
- 18 Changeux, 1985, p. 249.
- 19 Given that dendrites have non-linear properties, individual dendritic segments could be the brain's "basic computational units". Quartz and Sejnowski, 1997, p. 549.
- 20 Quartz and Sejnowski, 1997, p. 552.
- 21 For a clear analysis of these mechanisms in the visual cortex, see Katz and Shatz, 1996.
- 22 *Ibid.*, p. 550. Several interesting principles are used to explain the mature brain's functional organization; one is the so-called "geometric principle" through which information is collected in a topological way, spatially or conceptually related representations being realized in neighboring physical structures; the other is the "clustering" principle, through which related inputs onto dendritic segments result in a pattern of termination that mirrors the informational structure of the input, (*ibid.*, p. 549).
- 23 Katz and Shatz, 1996, p. 1137, Hurford et al. 1997, p. 567, Dehaene-Lambertz and Dehaene, 1997.
- 24 Some of the mechanisms allowing such a capacity to emerge from brain structure are described in programmatic terms in Quartz and Sejnowski, 1997.
- 25 On the biological importance of robustness, see Hammerstein et al. 2006.
- 26 See Proust, 2006b.
- 27 See Barsalou, 1999.
- 28 See Vygostky, 1978, Chap. 6.
- 29 See Sperber, 2006.
- 30 See Duncan, 2001 and Koechlin et al. 2003.
- 31 The stronger view of mental function that this article tries to elaborate does not deny that neuro-anatomical similarity across individuals and relative domain specificity in information processing make the modular idiom useful in certain respects. What it denies is that these modules are shaped by evolution as the conditions 1–3 above specify. If, as we argued above, the brain is developing in response to a changing environment, and function is determined by

developmental constraints and environmental demands, our definition of a mental function cannot mistake a given token or type of functional effect (the one that is observed given a social-historical developmental context), with the general function of adaptive control that it exemplifies in that particular dynamic context. This also suggests that interpreting psychopathology is more difficult than commonly acknowledged; the distinction between impairment and compensation, for example, is difficult to draw in the absence of an independent understanding of the mental function(s) (in this broad, dynamic sense of the term) that is supposed to be perturbed. I thank Dick Carter for critical observations associated to this point.

- 32 For the functional analysis of schizophrenia, as a good illustration of this point, see Proust, 2006a.
- 33 For a similar view on cognitive function, see Christensen and Tommasi, 2006.
- 34 These constraints are analyzed in more detail for their consequences on the architecture of the human mind in Proust, 2006b.
- 35 Decoupling is needed for representing an event at different times, representing fiction as not true, and understanding false beliefs. See Sterelny, 2003 and Proust, 2003.

Bibliography

- Amundson Ron and Lauder George V. (1994), "Nature's Purposes: Analyses of Function and Design in Biology", *Biology and Philosophy* 9 (4), pp. 443–470.
- Barsalou Larry W. (1999), "Perceptual Symbol Systems", *Behavioral and Brain Sciences* 22(4), pp. 577–660.
- Boyd Richard and Richerson Peter J. (1985), *Culture and the Evolutionary Process*, Chicago, The University of Chicago Press.
- Buller David J. (2005), *Adapting Minds. Evolutionary Psychology and the Persistent Quest for Human Nature*, Cambridge, Massachusetts, MIT Press.
- Changeux Jean-Pierre (1985), *Neuronal Man, The Biology of Mind*, New York, Pantheon Books.
- Changeux Jean-Pierre and Dehaene Stanislas (1989), "Neuronal Models of Cognitive Function", *Cognition* 33, pp. 63–109.
- Christensen Wayne D. and Tommasi Luca (2006), "Neuroscience in Context: The New Flagship of the Cognitive Sciences", *Biological Theory* 1(1), pp. 78–83.
- Christensen Wayne D. and Hooker Clifford Alan (2000), "An Interactivist-Constructivism Approach to Intelligence; Self-Directed Anticipative Learning", *Philosophical Psychology* 13(1), pp. 5–45.
- Cosmides Lea and Tooby John (1994), "Beyond Intuition and Instinct Blindness; Toward an Evolutionarily Rigorous Cognitive Science", *Cognition* 50, pp. 41–77.
- Dehaene-Lambertz Ghislaine and Dehaene Stanislas (1997), "In Defense of Learning by Selection: Neurobiological and Behavioral Evidence Revisited", *Behavioral and Brain Sciences* 20(4), pp. 560–561.
- Duncan John (2001), "An Adaptive Coding Model of Neural Function in Prefrontal Cortex", *Nature Neuroscience Reviews* 2, pp. 820–829.
- Edelman Gerald M. (1987), *Neural Darwinism: The Theory of Neuronal Group Selection*, New York, Basic Books.
- Etienne Koechlin, Chrystèle Ody and Frédérique Kouneiher (2003), "The Architecture of Cognitive Control in the Human Prefrontal Cortex", *Science* 302, pp. 1181–1185.
- Fodor Jerry (1983), *The Modularity of Mind*, Cambridge, Massachusetts, MIT Press.
- Hall Brian K. (1992), *Evolutionary Developmental Biology*, New York, Chapman & Hall.
- Jaenisch Rudolf and Bird Adrian (2003), "Epigenetic Regulation of Gene Expression: How the Genome Integrates Intrinsic and Environmental Signals", *Nature Genetics* 33, pp. 245–254.
- Griffiths Paul E. (1996), "Darwinism, Process Structuralism and Natural Kinds", *Philosophy of Science* 63, pp. 1–9.
- Griffiths Paul E. (1997), *What Emotions Really Are*, Chicago, The University of Chicago Press.

- Griffiths Paul E. (in press), "Evo-Devo Meets the Mind; Towards a Developmental Evolutionary Psychology", in Brandon R. and Sansom Roger (eds.), *Integrating Evolution and Development*, Cambridge, Cambridge University Press.
- Hammerstein Peter, Hager Edward H., Herz Andreas V.M. and Herzel Hanspeter (2006), "Robustness: A Key to Evolutionary Design", *Biological Theory* 1(1), pp. 90–93.
- Hauser Mark, Chomsky Noam and Fitch W. Tecumseh (2002), "The Language Faculty: What Is It, Who Has It, and How did It Evolve?" *Science* 298, pp. 1569–1579.
- Hurford James, Joseph Sam, Kirby Simon and Reid Alastair (1997), "Evolution might Select Constructivism". *Behavioral and Brain Sciences* 20(4), pp. 567–568.
- Karmiloff-Smith Annette (1992), *Beyond Modularity*, Cambridge, Massachusetts, MIT Press.
- Karmiloff-Smith Annette (1998), "Development Itself Is the Key to Understanding Developmental Disorders", *Trends in Cognitive Science* 2(10), pp. 389–398.
- Karmiloff-Smith Annette (2006), "Ontogeny, Genetics and Evolution: A Perspective from Developmental Cognitive Neuroscience", *Biological Theory* 1(1), pp. 44–51.
- Katz Lawrence C. and Shatz Carla J. (1996), "Synaptic Activity and the Construction of Cortical Circuits", *Science* 274:1133–1138.
- Maynard-Smith John (1983), *Evolution and the Theory of Games*, Cambridge, Cambridge University Press.
- Mayr Ernst (1961–1988), *Toward a New Philosophy of Biology*, Cambridge, Massachusetts, Harvard University Press.
- Millikan Ruth (1984), *Language, Thought and Other Biological Categories, New Foundations for Realism*, Cambridge, Massachusetts, MIT Press.
- Oyama Susan (1985–2001), *The Ontogeny of Information*, revised edition, Cambridge, Cambridge University Press.
- Panksepp Jaak and Jules B. (2000), "The Seven Sins of Evolutionary Psychology", *Evolution and Cognition* 6(2), pp. 108–131.
- Pinker Steven and Ray Jackendoff (2005), "The Faculty of Language: What Is Special about It?" *Cognition* 97, pp. 1–36.
- Proust Joëlle (1997), *Comment l'esprit vient aux bêtes, Essai sur la représentation*, Paris, Gallimard.
- Proust Joëlle (2003), "Can Radical Theories of Simulation Explain Mental Concept Acquisition?" in J. Dokic Jerome and Joelle Proust (eds.), *Simulation and Knowledge of Action*, Amsterdam, John Benjamins, pp. 201–228.
- Proust Joëlle (2006a), "Agency in Schizophrenics from a Control Theory Viewpoint", in W. Prinz and N. Sebanz (eds.), *Disorders of Volition*, Cambridge, Massachusetts, MIT Press, pp. 87–118.
- Proust Joëlle (2006b), "Rationality and Metacognition in Non-Human Animals", in S. Hurley and M. Nudds (eds.), *Rational Animals?*, Oxford, Oxford University Press.
- Proust Joëlle (in print), "Metacognition and Metarepresentation: Is a Self-Directed Theory of Mind a Precondition for Metacognition?", *Synthese*.
- Quartz Steven R. and Sejnowski Terry J. (1997), "The Neuronal Basis of Cognitive Development: A Constructivist Manifesto", *Behavioral and Brain Sciences* 20, pp. 537–596.
- Rizzolatti Guisepe and Arbib Michael A. (1998), "Language Within our Grasp", *Trends in Neurosciences* 21(5), pp. 188–194.
- Robert J. Scott, Hall Brian.K. and Olson Wendy M. (2001), "Bridging the Gap Between Developmental Systems and Evolutionary Developmental Biology", *BioEssays* 23, pp. 954–962.
- Roy Alice C. and Michael A. Arbib (2005), "The Syntactic Motor System", in K. Liebal, C. Müller and S. Pika (eds.), *Gestural Communication in Nonhuman and Human Primates*, Amsterdam, Benjamins, pp. 7–37.
- Shockley Kevin, Santana Marie-Vee and Fowler Carol A. (2003), "Mutual Interpersonal Postural Constraints are Involved in Cooperative Conversation", *Journal of Experimental Psychology* 29(2), pp. 326–332.

- Sperber Dan (2006), "Why a Deep Understanding of Cultural Evolution is Incompatible with Shallow Psychology", in N. Enfield and S. Levinson (eds.), *Roots of Human Sociality*, Oxford, Berg.
- Sterelny Kim (2000), "Development, Evolution and Adaptation", *Philosophy of Science* 67, pp. 369–387.
- Sterelny Kim and Griffiths Paul E. (1999), *Sex and Death, An Introduction to the Philosophy of Biology*, Chicago, The University of Chicago Press.
- Thelen Esther and Smith Linda B. (1994), *A Dynamic Systems Approach to the Development of Cognition and Action*, Cambridge, Massachusetts, MIT Press.
- Vygotsky Lev S. (1978), *Mind in Society. The Development of Higher Psychological Processes*, Cambridge, Massachusetts, Harvard University Press.
- Whiten Andrew (1997), "The Machiavellian Mindreader", in A. Whiten and R. Byrne, *Machiavellian Intelligence II, Extensions and Evaluations*, Cambridge, Cambridge University Press, pp. 144–173.
- Wimsatt William. C. (1986), "Developmental Constraints, Generative Entrenchment, and the Innate-Acquired Distinction", in W. Bechtel (ed.), *Integrating Scientific Disciplines*, Dordrecht, Martinus Nijhoff, 1986, pp. 185–208.

Philosophy of Cognitive Science

Daniel Andler

The rise of cognitive science in the last half-century has been accompanied by a considerable amount of philosophical activity. No other area within analytic philosophy in the second half of that period has attracted more attention or produced more publications. Philosophical work relevant to cognitive science has become a sprawling field (extending beyond analytic philosophy) which no one can fully master, although some try and keep abreast of the philosophical literature and of the essential scientific developments. Due to the particular nature of its subject, it touches on a multitude of distinct special branches in philosophy and in science. It has also become quite a difficult, complicated and technical field, to the point of being nearly impenetrable for philosophers or scientists coming from other fields or traditions. Finally, it is contentious: Cognitive science is far from having reached stability, it is still widely regarded with suspicion, philosophers working within its confine have sharp disagreements amongst themselves, and philosophers standing outside, especially (but not only) of non-analytic persuasion, are often inclined to see both cognitive science and its accompanying philosophy as more or less confused or even deeply flawed.

The sensible way to go under the circumstances, or so one might judge, would be to pick a sample of salient topics, in the present case, philosophical discussions of some central foundational issues, in the hope thereby of giving the reader a sense of what the field is about. This however is not the path I propose to take. There are two reasons for choosing another tack. The negative reason is that there is now available a plethora of excellent expositions, of any length one might desire, from one-page summaries to chapter- or volume-length introductions, of central topics in philosophy of mind (which constitutes in turn the core of what most philosophers think of as philosophy of cognitive science: more on this in a moment).¹ Producing one more such exposition seems hardly worth the effort. The positive reason is that philosophy of science in general has a number of goals not all of which consist in

D. Andler (✉)

Département d'études cognitives

École normale supérieure

29 rue d'Ulm

75005 Paris, France

E-mail: daniel.andler@ens.fr

elucidating foundational issues; for example, there are issues of methodology; there are conceptual problems linked to empirical issues which seem not yet ripe for direct scientific resolution by available means. But there is also the more general concern of providing a perspective on the structure and dynamics of a field, its relations to other areas of inquiry, its purported limitations or misconceptions, its future directions. This applies to science as a whole as it does to the specific fields and disciplines, however broad or narrow. And it would seem to apply all the more to fields which have emerged recently, and which therefore give rise to questions about where they fit in the overall scheme of inquiry, why they didn't appear sooner and whether they are here to last. In short, philosophy of cognitive science can also, and perhaps should, be thought of as a division within the philosophy of science, on par with philosophy of biology, philosophy of economics, etc. This may seem obvious, but it is not how it is usually treated.

This at any rate is how the present chapter proposes to view the topic. In broadening its scope, it will necessarily sacrifice depth, but will still stop short of providing answers to all the concerns just listed. I will begin with a discussion of metatheoretical issues, which will prepare us for the sequel. I will not attempt to draw the contours of cognitive science. This would go without saying in the case of any other field: who would expect a 'definition' of mathematics in a chapter devoted to contemporary philosophy of mathematics? Cognitive science is different: its image is blurred. It will be less so, one may hope, by the time the chapter closes. Meanwhile however, it might be helpful to some readers to have a working definition. Let's content ourselves with the following: Cognitive science corrals a variety of disciplines and approaches with the aim of providing an integrated scientific account of the mind, its states, processes and functions. If 'mind' is thought to command a premature commitment to a dubious ontology, one can for the time being substitute 'behavior', withholding any strong preconception about what counts as behavior and what may or may not enter in the sought-after accounts. Finally, the reader who has trouble seeing what distinguishes cognitive science thus characterized from (scientific) psychology is asked to only accept the following amendment: cognitive science is, as it were, psychology pursued by novel means; it draws on any potentially relevant discipline (the main contenders being neuroscience, computer science and related modeling techniques from physics and mathematics, linguistics, philosophy and parts of social science), and its detailed agenda is thereby broadened far beyond the ones pursued in previous epochs of psychology.

1 Metatheoretical Issues

1.1 *The French Dimension*

Philosophers in France working on cognitive science are for the most part wedded to an internationalist view of scholarship. For them, the very idea of a 'French

cognitive science' or a 'French philosophy of cognitive science' is reminiscent of sorry episodes in the history of science (German physics, Soviet genetics,...) and, more to the point, of a recent period where French academia, especially in the humanities and sciences of man, was isolated from the international community and entered a phase of parochialism.

However commendable, these internationalist convictions should not lead one to take it as a necessary truth that there is nothing to be said about cognitive science, or about philosophy of cognitive science *in* France. It certainly makes sense that there would exist local schools in those areas, as there are in every enduring academic field.² It is neither absurd, nor a political fault, to inquire whether there are, or have been, specific traditions which originated or were developed in France. The answer turns out to be mixed, and cannot be expounded in any detail here.³ As for cognitive science proper, the prominent American psychologist David Premack was fond of saying that it was 'invented' by the French Nobel laureate Jacques Monod, who founded the Centre Royaumont pour une science de l'homme, a think-tank *cum* conference center which hosted in 1975 a memorable encounter between Noam Chomsky and Jean Piaget.⁴ Even interpreted with a grain of salt, this view can be taken as a conclusive refutation of the popular conception of cognitive science as a US import.⁵ France had strong traditions in the neurosciences, in linguistics, in anthropology, in mathematics, in cybernetics. There were also a small number of scientifically-minded psychologists spread over France and neighboring French-speaking countries.⁶ These resources could be pooled to constitute small informal groups, within which a culture emerged which we can retrospectively identify as cognitive-scientific, and on which the present generation was able to build in its successful efforts to set up cognitive science in the contemporary French academic scene.

As for the philosophy of cognitive science, its fate has been closely connected to the development of analytic philosophy, which was all but barred from France after World War Two. Only in the late 1960s were a small number of young philosophers able to cultivate it, and it took another 20 years or so for analytic philosophy to gain acceptance. On that front, then, there is no denying that the now well-established group of analytic philosophers of cognitive science working in France, some of whom enjoy an international reputation, are true cosmopolitans. But next to them, there are a small number of philosophers who lean on Kantianism or on phenomenology (mostly in Husserl's and Merleau-Ponty's traditions) to approach cognition from an angle unfamiliar to most analytic philosophers. The mathematician René Thom's visionary ideas in natural philosophy and in theoretical biology have also had some influence, and so have some successful non-technical books written by prominent neuroscientists.⁷ These strands are more clearly French, but limitations of space would force me to give their ideas short shrift, or else not present the international context in which they are deployed, something they would disapprove of for the reasons stated above.⁸

The upshot is that for the most part this chapter will dwell on the philosophy of cognitive science *simpliciter*, without an eye on a specifically French source or style.

1.2 *Philosophy of/and/as/in Cognitive Science: A Logical Geography*

Up to now I have been using ‘philosophy of cognitive science’ as the most general term embracing all forms of philosophical activity connected in one or another way to cognitive science, but I indicated that there exist under that general rubric some rather different forms of inquiry. Indeed, there is no agreement on names for these various forms, nor on the significance of the differences. For some, there is no reason to draw sharp distinctions, or any distinctions at all, between philosophy of psychology, philosophical psychology, philosophy of mind, philosophy of cognitive science or cognitive philosophy.⁹ But the fact that there are indeed no crisp boundaries, that some issues can be seen as taking part in different projects, and that there exist multiple connections between topics belonging to different areas, does not imply that distinctions cannot be usefully drawn. In fact, I claim that such distinctions are an integral part of the preliminary agenda of the philosophy of cognitive science in the wide sense, one which hasn’t received sufficient attention.

There are, as I see it, two dimensions of contrast to consider. Along the first axis, one can plot proximity to (cognitive) science. Proximity involves either collaboration, or sympathy, or both. Near one end of the line, one finds research programs in which philosophers and scientists from one or another discipline, or sometimes several, attempt to provide a solution to some specific problem concerning a cognitive phenomenon. This effort, when successful, leads to a scientific achievement which both owes to, and rewards, the philosophical investment which went into it. This is the sense in which philosophy is one of the basic components of the cognitive federation: the relation is one of inclusion, philosophy *in* cognitive science (or again, philosophy *as* cognitive science). Examples abound in such areas of study as visual perception, reasoning, linguistic communication, numerical knowledge, voluntary movement, ‘mind-reading’, social skills, to pick just a few examples among dozens. Such shoulder-to-shoulder activity makes sense only at the cost of renouncing the traditional view of the philosopher as a respectful witness, an expositor or again an appraiser, of science. It also implies the maximal degree of sympathy, *viz.* direct involvement.

Near the other end of the line, one finds the standard situation where philosophy, allied with history, examines and appraises cognitive science as an enterprise situated at some distance, somewhat like philosophy of art stands (for the most part) outside art, or general philosophy of science stands outside science (by necessity, as there is no such thing as ‘general science’), or philosophy of chemistry, in all but exceptional cases, stands outside chemistry. Philosophers who operate in such a framework typically do not attempt to directly contribute to the enterprise they are appraising. Sympathy may be present, but with a certain professional distance; and it may also be altogether absent, when the philosopher finds the science (or the art form) she is examining flawed and develops an (informed) critique. This type of inquiry is, in the case at hand, philosophy *of* cognitive science in the traditional, restricted sense.

There is a continuum, rather than two discrete positions, along this proximity/sympathy axis, and this is important. As one leaves the high-proximity end (philosophy in/as cognitive science), the problems become increasingly conceptual, the philosophical component takes precedence and one moves away from the day-to-day scientific work. One gets nearer the traditional position of appraisal and concomitant philosophical speculation, and one enters the area of ‘ontological problems’.

The second axis measures distance from psychology in the traditional sense. At the near end, there is psychology itself; at the far end, cognitive science in its widest sense, where psychology as traditionally construed no longer occupies a privileged, central position. Some readers may be surprised to learn of the existence of such a conception, but the history of cognitive science began with claims by artificial intelligence (AI) to subsume and thus supersede psychology, and has now come to a point where similar claims are made on behalf of cognitive neuroscience. One of the main duties of philosophy of cognitive science is to critically examine and compare these opposing views of the essential nature of the field. But notice again that there is a continuum: on some views, psychology retains a distinguished position without constituting the very heart, let alone the entirety, of the field.

We are now in a position to draw a map of the area (see Fig.1). We start by pretending there is a discrete 2-dimensional logical space with four positions. Horizontally, we have FAR / CLOSE with respect to (cognitive) science. Vertically we have CLOSE / FAR with respect to traditional psychology. Then in the lower left position we find philosophy of psychology; above, philosophy of cognitive science (in the strict sense); in the lower right there is philosophical psychology; in the upper right, philosophical cognitive science (or philosophy as cognitive science).

Finally, we need to introduce two complications in order to get a realistic picture with familiar labels. The first is to replace the discrete positions by a continuum along both dimensions. This creates a middle zone straddling all four positions. The only available label for this middle zone is philosophy of mind. It intersects with philosophical psychology and philosophical cognitive science, on the right, and with philosophy of psychology and philosophy of cognitive science, on the left. However, philosophy of mind extends beyond the entire space, as some philosophers working on the mind raise metaphysical issues quite independently of any science of the mind, past, present or future. Thus there is a part of philosophy of mind which resolutely straddles both cognitive science and its philosophy (this part is sometimes called cognitive philosophy, or again philosophy of cognition), and a (admittedly smaller and less visible) part which is light-years away from it all, with all degrees in between. Similarly, one may be tempted to take philosophy of psychology to be a proper part of philosophy of cognitive science, and philosophical psychology a proper part of philosophical cognitive science. But that would be to ignore or preclude the possibility of a non-cognitive form of psychology, together with a philosophy of non-cognitive psychology, or rather, a philosophical examination of the claim that psychology is not, and should never become, entirely immersed in cognitive science.

Some readers, especially among cognitive philosophers,¹⁰ might object to this taxonomy, finding it otiose or at least unnecessarily complicated (or worse yet:

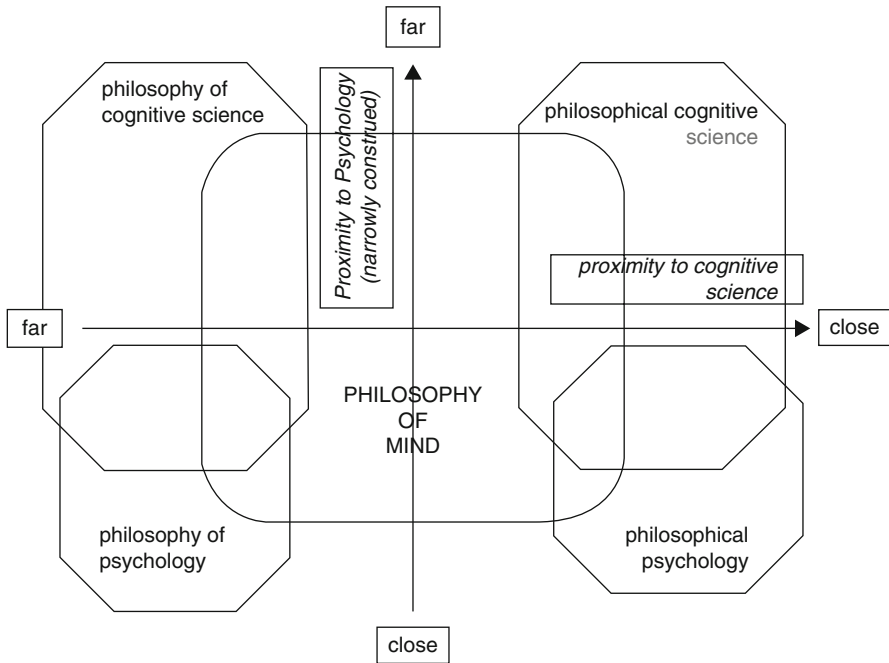


Fig. 1 See text

making a fuss over labels). Yet, on the one hand, it seems important to emphasize the co-existence of rather different research programs involving philosophy and cognitive science. On the other hand, it seems no less crucial to allow for philosophical or scientific enterprises which are both concerned with the mind and free of any analytical connection to cognitive science, although clearly, as their distance from cognitive science grows, the relevance of these enterprises for philosophy of cognitive science in the wide sense vanishes to zero.

1.3 Philosophical Styles and the Place of History

As hinted above, philosophy of cognitive science (*lato sensu*) and especially philosophy of mind (in the restricted sense of cognitive philosophy) have been so deeply linked to the development of analytic philosophy in the last quarter century that up until recently there have been few contributions from other traditions in philosophy, such as phenomenology. There were exceptions: AI, an early avatar of cognitive science, was critically examined in the 1970s with the help of tools drawn mostly from existential phenomenology or hermeneutics.¹¹ This enterprise was however conducted in North America and in the framework and style of analytic philosophy. There is now a growing body of work which blends the analytic style with phenomenological themes.¹² On the other hand, philosophers trained outside the analytic

tradition, which compose, for example, the vast majority of French philosophy, have by and large remained unconcerned by or very dubious about both cognitive science and the associated philosophical inquiries.¹³ Perhaps these differences will become less relevant, as parts of analytic philosophy become more permeable to ‘continental’ influence, and vice-versa: opinions vary widely on the probability and on the desirability of such a *rapprochement*. But it would take us too far afield to discuss the infamous ‘analytic-continental divide’ and its repercussions on the topic at hand.

Let us turn our attention, instead, to a more restricted issue, that of history of science. Continental, and more particularly French, philosophy of science, is wedded to history, and analytic philosophy of science today, or what is sometimes called ‘post-positivist’ philosophy of science, has espoused the view that philosophy of science cannot dispense with a historical perspective. Thus philosophy of cognitive science (*stricto sensu*) would seem to go hand in hand with history of cognitive science. The trouble is that this latter field remains to this day quite underdeveloped, leaving the former somewhat handicapped. In the present chapter, some minimal historical landmarks for cognitive science will be provided in passing, yet nothing will be said about the history of philosophy of mind. Although this conforms to the usual treatment of the topic, and to accepted practice within analytic philosophy, it must be acknowledged as a deficiency. First, there is an inconsistency in providing some historical perspective, however scant, for the scientific part, narrowly construed, of cognitive science, while denying the more centrally philosophical part a similar treatment, given that the multiple connections between the two parts is a key feature of the field. Second, it is quite likely that an account of the genesis of the main ideas and concepts in philosophy of mind, both currently in the mainstream and heterodox or less fashionable, would throw some light on the present debates and provide conjectures on the underlying dynamics; it would assuredly also considerably help the non-specialists find their way in the thicket of the existing literature. Unfortunately, not only do I lack the expertise to provide this much needed historical perspective; but it does not seem readily available.¹⁴ To be sure, there are collections which include a few classical texts, from Plato to Russell.¹⁵ However, they suffer from two unavoidable shortcomings: the excerpts are perforce taken out of context, and invite a whiggish interpretation by the historically innocent; and they carefully avoid providing a narrative, without which, however subjective and incomplete, not much can be gained, beyond encyclopedic knowledge which is of little use in achieving a true grasp of the current situation. In addition, these collections tend to favor the distant past (sometimes going back to Plato, Aristotle, more often starting with Descartes, carefully avoiding Kant and reaching for Brentano), and then jump to contemporary authors, leaving not much room for recent authors (*e.g.* Wilfrid Sellars, to name just one recently resurrected figure, Gilbert Ryle or Herbert Feigl, or again the Pragmatists), who are likely to have directly inspired some of the more esoteric proposals under scrutiny today, and almost completely neglecting non-naturalist thinkers, followers of Husserl, Merleau-Ponty or Wittgenstein, not to mention Ernst Cassirer or Suzanne Langer.

There are several well-known retorts to these historical scruples. The first is that they are misplaced, either because philosophy of mind is an essentially new enterprise which is no more interestingly related to its distant ancestry as say contemporary physics is to Archimedes; or because the genesis of a philosophical problematic is at best of tangential help in our attempts to clarify and solve the problems as they are set up today. The second is that the dream of a systematic philosophy, which seems to underlie this perhaps unreasonable longing for history, belongs to the past, or else must be thought of as a blessing rather than a goal: one cannot aim for it, one can only hope that it be realized, at certain times and within certain communities or single thinkers, as a form of culture or wisdom. The third is that historical studies directly relevant to present concerns are not truly feasible, be it for conceptual or institutional reasons.¹⁶ But whatever the merits of these responses, they are ineffective as cures against my scruples: I still deplore the absence of an informative historical frame for philosophy of mind.

We are finally ready to examine some substantive issues. The rest of the paper comprises two sections. Sect. 2 is devoted to the central task which philosophy of mind has set itself, *viz.* to provide cognitive science with a conceptual framework. Sect. 3 concerns the fit between the framework and cognitive science, and includes discussions of a sample of issues internal to the field.

2 A Conceptual Framework for Cognitive Science

Cognitive science is often simply defined as the science of the mind, while philosophy of mind can be seen, first and foremost, as the exploration of the ontology of the mind. As we limit ourselves, from now on, to that part of philosophy of mind which is in direct contact with cognitive science (cognitive philosophy, or philosophy of cognition), we can view its general aim as providing a foundation to cognitive science. Such a foundation is to be sought neither in pure *a priori*, conceptual analysis, nor in some kind of inductive generalization from the practice of cognitive science. It consists rather in searching for a reflective equilibrium between the ontological principles suggested by philosophical inquiry and what may be called the ontological practices, or perhaps the implicit ontological commitments of cognitive science. Paraphrasing a famous title,¹⁷ we are asking, as it were, What is a Mind, that Cognitive Science May Know It, and Cognitive Science, that It May Know a Mind? While the second part of the question falls squarely in the province of the philosophy of cognitive science, the first states in the most general way the purpose of the philosophy of mind. Thus combined, they are seen as mutually dependent. One must however choose a starting point, and I shall begin with the conceptual framework which philosophy of mind has proposed for cognitive science. The focus will be on what is often called the 'classical' paradigm, but some mention will be made of possible deviations, and what will be offered is a liberalized form of the paradigm, which I believe to be at this

point in time the inevitable point of departure for any inquiry into cognitive science (I will return to this point in Sect. 3).

2.1 The Mind-Body Problem, Physicalism, Functionalism

A survey of the mind-body problem, however compressed, cannot be attempted here, for reasons of space.¹⁸ Fortunately, in the somewhat limited perspective I have just proposed, such a survey is not strictly necessary. What we need to focus on are the points of contact between the philosophical analysis of the problem and the scientific practice.

The first thing to notice is that the connection is not as strong as one might think.

Although there are many variants of the mind-body problem, they share a core, which is a longing for an understanding of the way in which the mental realm, which appears to float free of the physical realm, might fit, or, alternatively, of the reasons why it could not conceivably fit, in an overall picture which accommodates everything that, and nothing but what exists in the natural world. Thus one might think that the analysis of the mind-body problem which philosophy of mind is supposed to deliver is crucial to cognitive science, in the sense of being an ‘enabling’ factor. Or, conversely, that cognitive science provides philosophy of mind with essential empirical, or more broadly, scientific ingredients of, or constraints on a solution to its central problem.

The starting point for this way of thinking is the core belief, accepted by many philosophers, and scientists, that cognitive science will inevitably provide in due course a thoroughly naturalistic account of the mind, on par with our current understanding of lightning, eclipses, tropical storms or illness which were once held to elude explanation by natural causes. This core belief in turn is supported by two lines of thought. One, the ‘fast track’, enlists, severally or jointly, two arguments. The first takes as premise the perfect record of physics (construed in a sufficiently broad sense as to include chemistry and all the ‘special’ natural sciences) in accounting for any and all aspects of the real world which it has attacked, and proceeds inductively to the conclusion that the mental realm is likely to succumb as well. The second takes as premise the success of evolutionary theory in accounting in strictly natural terms for the presence of complex functional systems in the living world, and proceeds inductively to the conclusion that the mind, a complex functional system if there ever was one, can and eventually will be seen as nothing over and above an evolved, and hence natural part of the living world.

The second line of thought (the ‘slow track’) starts with a consideration of the ongoing work in cognitive science, which began, in recognizable form, in the 1950s, and is generally regarded as flourishing,¹⁹ unimpeded by any trace of a major crisis which, as happened for the first phase of artificial intelligence, might spell the end of the enterprise. To the contrary, the rise of cognitive neuroscience, fuelled by functional imagery, seems to both considerably enrich the toolkit of cognitive sci-

ence and reinforce its character as a *natural* science. Thus cognitive science is in the process of showing the natural character of the mind by actually proceeding, step by step, to an effective naturalization of the mental realm.

Both tracks converge on the *philosophical* thesis of naturalism, which may or may not be further specified as physicalism. The thesis, in both its liberal and strict (physicalist) versions, generates in turn a budget of philosophical puzzles –such as the apparent impossibility of mental causation due to the alleged causal closure of the physical– which constitute an important part of the agenda of philosophy of mind.

However, the general arguments briefly sketched above in favor of naturalism, or more stringently, physicalism, are notoriously inconclusive in the eyes of those not already persuaded, and in some sense they seem indeed to beg the question. In fact, even the keenest defender of philosophical naturalism can see that a full naturalization of the mind delivered by cognitive science remains a distant prospect. But the discussion lies for the most part outside the immediate agenda of the philosophy of cognitive science, except for the methodological examination of the naturalization programs deployed in the field, which will not be undertaken here due to space limitations. The important point here is that cognitive science does not require a prior belief in the inevitability of a complete success of naturalization. At most it needs reassurance that there is no incontestable proof or overwhelming evidence arising from other areas of science, including the formal sciences,²⁰ or from non-scientific sources, which would establish beyond reasonable doubt that the mind (or at least some essential dimension of it) is not accessible to science.

This being said, cognitive science and philosophy of mind do lend one another considerable support, as indicated, and for the general reason given at the outset. The apparent conflict arises from the mind-body problem. How can the philosophical discussions of this issue both be and not be relevant for cognitive science, and conversely how can the general orientations and results of cognitive science both be and not be relevant for the resolution of the philosophical mind-body problem? The short answer is that philosophy suggests, and cognitive science supports, precisely the idea that a solution of the philosophical problem, at least in the traditional terms of dualism *versus* monism, is not required for cognitive science. The detailed answer consists in an exposition of the doctrine of functionalism.

Functionalism means somewhat different things for different authors at different times,²¹ but the core of the doctrine is that mental entities such as beliefs and desires, pains and rememberings, regrets and fears, are functionally defined kinds of inner states which can be entirely individuated by the role they play in the dynamics of the cognitive system, with sensory stimulations and motor responses constituting a set of observable boundary conditions. So that believing that the sun is setting is individuated by the relations which obtain between that belief, other beliefs, desires, perceptual states and motor responses: the belief that the sun exists, the perception of the sun nearing the horizon, the knowledge that the air cools at sunset, the desire to keep warm, the memory of where the sweaters are stored, the motor commands leading to the cupboard, etc. The crucial point is that there need not be anything further to know, or perhaps even to be known, about desires and other mental states.

We can remain blissfully ignorant of their ontological status, of the stuff they are cut out from, whether material or ethereal, and go about uncovering the laws of thought, or in other words, conduct the business of empirical psychology.²²

This core idea calls for specification along a number of dimensions. The more fully worked-out theories divide up in three main types. *Commonsense* (or *analytic functionalism*) holds that mental states are actually *defined* by the set of ‘platitudes’ or commonsense regularities in which they are unreflectively seen to enter: the *meaning* of ‘belief (B) that aspirin relieves headaches’ is *exhaustively* provided (albeit implicitly) by the myriad generalizations regarding headache-relieving aspirin episodes (such as: Having a headache and having belief B tends to induce, *ceteris paribus*, aspirin-absorbing behavior). These are platitudes insofar as they are analytic truths, holding in virtue of the meaning of the mental state terms involved. Someone unable to immediately see them as true is not a psychological idiot, but simply fails to grasp the meaning of these terms.

Psychofunctionalism (or *empirical functionalism*) rests on a distinction between the *meaning* and the *reference* of mental state terms. What we need to know, regarding, say, the generic notion of belief, or one particular belief such as B above, is for science to discover. Following recent discussions in philosophy of language, many philosophers see the meaning of a natural-kind term like water as fixed by a combination of everyday linguistic and other social practices, while the reference of water (stuff made of H₂O molecules) is for chemistry to determine. Psychofunctionalists extend this view to mental state terms. Psychology, embedded in cognitive science, will ideally determine the causal nexus characteristic of any given mental state (or process).²³

Machine functionalism is Hilary Putnam’s initial version of functionalism,²⁴ and it equates mental states with the states of a Turing machine or more broadly a computational system. Despite the fact that it has been all but abandoned in its original form, also by its proponent himself,²⁵ machine functionalism is by far the most relevant for cognitive science as what provides it with a (partial and provisional) foundation, for it is strongly connected to the ‘computer model of the mind’, or more accurately, to the computational theory of mind.²⁶

2.2 *The Computational Theory of Mind (CTM)*

A natural and quick way to introduce the CTM is to first consider another reason why machine functionalism is so important in this context: it provides a vivid, accessible illustration of the main conceptual features of all forms of functionalism.

The most central of these is that functionalism allows for *multiple realizability*. Chairs, functionally defined as pieces of furniture providing support for sitting, can have all sorts of shapes, be constructed in all sorts of ways, out of all sorts of materials. Similarly, a lever, a pulley, a wheel, a carburetor, a corkscrew are what they are not by virtue of their *specific* or *intrinsic* properties, but by virtue of their ability to fill a certain role in the framework of a larger system. Functional-

ism as a doctrine of the mind likewise views mental states and processes as roles variously filled by (human) brain states and processes (different brains, or even the same brain at different moments, filling the roles in different ways), and also conceivably by certain non-human brains, and even by artefacts such as computers or other complex devices. Computers, or their notional paradigm the Turing machine, provide a perfectly straightforward example of multiple realizability: 'believing' that 3 times 7 is 21 is 'realized' in an indefinite variety of ways according to the program (or machine table) which executes an algorithm for multiplication; and what makes it the 'belief' which it is is entirely, and non-mysteriously, a matter of the causal links between that state and other states and processes. The very same 'belief' can be present in computers with very different logical and material structures.

The computational theory of mind originates in this intuition, to which Putnam's papers in the mid-70s only gave a philosophically arresting form, as it had been adumbrated by Turing himself in his 1950 *Mind* paper, and developed by the fathers of artificial intelligence from the mid-1950s onward.

There is another crucial dimension to Turing's idea: computation as he (re)defined it is mechanical, hence poised for inclusion in the physical realm. The mind can be seen as a set of processes operating on a set of states. The processes are natural to the extent that they are causal, and if computation is mechanical (in a conceptual sense), Turing's work shows, and modern computers prove, that the mind is actually mechanizable, *i.e.* realized by a concrete mechanical system. What appears to some as the beauty of Turing's proposal (and to others as a weakness) is that it allows, yet does not force, a materialist solution.

But now we clearly see that the computational intuition is nonetheless not sufficient to generate a theory of the mind.²⁷ This it achieves only when grafted onto a much older idea, which originates in the philosophical psychology of the 17th century and permeates scientific psychology until the advent of CTM's predecessor, behaviorism. This venerable idea is that of a (mental) *representation* (or *idea* in the language of Descartes or Locke). Computation operates on *data*, in the original sense, or *inputs*. Now data *need* not represent (stand for) anything beyond themselves as material formal entities. The essence of computation is made (almost) entirely manifest by an example such as: concatenating XXX with XX yields XXXXX, where each one of the five Xs, as a material token, provides the computational device all it needs to proceed.²⁸ However, the most salient attribute of the mind is its ability to deal with entities lying clearly outside itself and forming an open-ended collection. In simple terms, it is constitutive of a mind, under nearly any construal, to be engaged, via the various perceptual modalities, in continuous interaction with the world (including other mind-bearing organisms). To determine what distinguishes minds in perceptual contact with the world among systems causally affected by the world is a problem of the highest difficulty, and the concept of a mental representation is offered as the starting point, indeed as the lynchpin of a possible solution. When an asteroid hits the Earth, the result is a crater. When a rock impinges on my visual system, a representation of the rock is formed in or by my mind. Although perception is presumably not the only process by which

representations are generated, it is enough to make *some* notion of representation very nearly impossible to dispense with.

To summarize: mental states are representations, appropriately labeled according to their status as propositional attitudes (beliefs, desires, fears, regrets...), and mental processes driving the mental dynamics, the transitions from state to state, are computational.

Computation and representation can now be assembled into a skeletal model of the mind. The connection is located at one precise point: the formal tokens on which the computation operates now represent aspects of the outside world; in fact, they can represent anything at all, including facts and events concerning the model itself, and including non-facts and non-events involving real or unreal entities. The representational content of a token X is dubbed *information*: it is, in roughly the usual sense of the word, information *about* what X is a representation of. This is why the CTM is more properly called by some authors the ‘computational-representational’, or ‘computational-informational’ theory of the mind (and, for the sake of completeness, let me repeat that sometimes ‘functionalism’ is used as yet another synonym).

As I have introduced it, CTM requires quite a bit of building up to start looking like something more than a ghost of a theory, a ‘we-know-not-what’.²⁹ This is the purpose of the next two subsections.

2.3 *Rationality, the Systematic Mind, and the Language of Thought Hypothesis*

The intellectual landscape in which the CTM emerged was shaped by the accomplishments of logic in the 1930s. Of prime importance for CTM is the notion of a formal system. Formal systems were adumbrated, though not fully constructed, by Frege, who was after a ‘good’ language for mathematics and science (this was of course, an old philosophical goal, but Frege was aiming this time for a scientific solution). Formal systems are exactly what CTM requires as a medium of computation *and* of representation. Their generative structure endow them with representative power. As Leibniz had dreamt, and Frege shown possible, they potentially contain (*i.e.* they can generate) symbols for any number of entities and states of affairs (under an essential proviso, which will be discussed presently). And their compositional nature makes them suitable for computation. This in turn involves two crucial properties: mechanical semantics and mechanical inference. Let us spell this out.

Leibniz’s dream and Frege’s aim was a scientific language in which the pursuit of truth could be conducted in the fashion of elementary arithmetic: ‘Let us calculate’. Frege turned logic, which for Aristotle was the canon of valid inference, into a language, without sacrificing its original function. Aristotle’s notion of formality contains the modern idea of syntax achieving the right semantic effect. In the syllogism ‘All As are Bs, some Cs are As, therefore some Cs are Bs’, the mind remains on the path

of truth by merely noticing the correct formal identities (as we would say, correctly identifying the types of the various tokens): it need not know what A, B, C stand for, let alone investigate the C population to ascertain whether some are also Bs. In other words, rationality (construed as truth-following) is achieved by formal or syntactic means only. But there is no *operation* involved: the inference rests on a *noticing*. In the modern (Hilbertian) perspective, an inference is an operation, notionally involving the manipulation of marks on paper. The second modern ingredient missing in ancient logic is the ability to encode or represent relations: only monadic predicates have a formal expression. Frege (and Peirce) revolutionized logic by introducing binary predicates and the ontological operators, quantifiers, which they require. They paved the way for the definition, in the 1920s, of a formal language in which all of the features of a situation in a given domain can be expressed, and where the syntax (the set of formal operations) ‘mimics’ the semantics. Gödel and Turing took the last essential step by showing that the formal operations are computable, and thus mechanizable.

The entire paragraph above, it is important to notice, has nothing directly to do with psychology, a theory of the mind. It has to do with thought, which is what a mind does. There is no simple abduction from what the mind does to how the mind is constituted: extra work, and considerable scientific imagination, is required. In fact, for centuries it seemed that the fact that the mind could apparently do just about anything meant that it could not have much structure.

The language of thought hypothesis (LoTH) is precisely the conjecture that the mind itself, regardless of what it is doing, in the familiar sense, at any given time, operates like an automated formal system, *i.e.* applies inference rules to formulas of a formal language, called ‘mentalese’ or ‘language of thought’ (LoT). LoTH presupposes CTM, and is often in fact taken to be a way of specifying it.

This way of putting LoTH is however less than fully satisfactory for two reasons at least. The first is that having the mind ‘apply inference rules’ seems to force one into either an infinite regress or a counterintuitive psychology. For if we mean to say that the mind applies rules in the same sense as we would say, for example, that a child adds 12 and 17, or a shopkeeper returns change, then we want to know again how the mind accomplishes this task (aren’t we after an account of how a child adds and a shopkeeper returns change?). Alternatively, if what we mean is that, contrary to what appears on casual introspection, whatever we are accomplishing mentally (thinking of how to begin a lecture, remembering a deceased parent, spelling our name on a form, making our way to our airport gate, breaking up with a lover, grasping a missing link in a story, etc.) *we* are actually applying rules to formulas, then we have to face severe objections of an empirical or phenomenological kind. But *neither one or the other is what is meant or implied by LoTH*. Rather, the suggestion is that when we accomplish a typical mental task, there are components of our cognitive apparatus, *i.e.* our central nervous system, whose trajectory is conspicuously described as the application of some rules to some formulas of an inner language. In a nutshell, when *I* accomplish any one among the myriad mental tasks people go through as a matter of course, the *cognitive system* responsible for my having a mind, and the mind I have, applies rules to formulas. The fate of LoTH, and probably of CTM as well, hangs on the possibility of making this suggestion fully intelligible.

The second reason why the simple formulation above is unsatisfactory is that it runs together two kinds of computational process. One is composition and regards *representations*, which are typically names for things or for states of affairs. The name, in mentalese, for Nicole's grandmother (call it X) is made up of the mentalese name N for Nicole, the mentalese word G for grandmother, and the mentalese word P indicating the relation of possession; and the process by which the mind goes from the set of symbols {N, G, P} to X is computational. Likewise, the fact that the cat is on the mat is represented by a formula of mentalese –something like $On_m \text{ Cat}_m \text{ Mat}_m$ – which is the result of a computational process acting on its components On_m , Cat_m , Mat_m . The other kind of computational process concerns *inferences*, which are the transformations undergone by sets of representations. For example, from the facts that (a) the cat is on the mat and (b) the mat is in the back porch, there follows (c) the cat is in the back porch. How I get to believe (c) results from *my cognitive system* having applied a rule of spatial logic to the mentalese counterparts of (a) and (b). (For readers suspicious of spatial logic, the example can be recast as (a), (b') If something is on the mat, it is in the back porch, (c)). Now it is just as important that the process of getting from the representation of some entities or relations to the representation of either a composite entity or a state of affairs be computational as the property of getting from some premises to a valid conclusion be computational, for the mind requires both kinds of moves to get on with its business.

The point of going over the familiar genesis of the modern idea of a formal system is to stress (i) the non-trivial character of this invention; (ii) the non-obvious (indeed, controversial) inference from the notion of a formal system as a medium *on which* the (logical) mind operates to the hypothesis of a formal system as a medium *by which* the mind, engaged in whatever task, operates; and therefore (iii) the unavailability of LoTH and hence of a reasonably worked-out form of CTM to thinkers working prior to the development of modern logic. I think all three of these points, however unoriginal, deserve to be made, as they are often not taken in, a cause for unnecessarily long expositions of objections which miss the target.³⁰

Finally, two arguments in favor of LoTH must be mentioned. The first counts more generally in favor of CTM. I said earlier that a prime feature of the mind is that it seems able to carry out just about anything, from determining the 111th prime number to imagining a purple cow with wings, buying bread or writing a haiku, and that this argued against any account of the mind which would give it any determinate structure. Now if the mind is somewhat like a computer, which is what LoTH implies, then it could be an approximation to a *universal* Turing machine, *i.e.* a device which can effect *any* computation.³¹ Together with the representational potential of formal languages, this universality plausibly confers a high amount of versatility to the system.

The second argument, developed at length by Fodor and Pylyshyn, consists in showing that LoTH is the best explanation (in fact the only available explanation) for some central features of the set of thinkable thoughts, *viz.* its productivity (the fact that for any finite set of thinkable thoughts, there is a thinkable thought which doesn't belong to it) and its systematicity (in the sense given by Fodor: it leaves no gaps, so that if 'John hates Peter' is thinkable, so is 'Peter hates John', 'Peter hates

someone', etc.). *If* thinking consists in moving between structured symbolic representations, then the productivity and systematicity of the set of thinkable thoughts is nicely accounted for by LoTH.³²

2.4 Intentionality

The general problem of intentionality is often seen as the central problem in the philosophy of mind, and it consists in accounting in ontologically acceptable terms for the *aboutness* relation which holds between a mental event such as a thought or a belief and what that thought or belief is about. There seems to exist nothing else in the world which has this essentially directional character. It does not do to say that a thought of mine is about the Eiffel tower just because it is associated, in however strong or privileged a way, to the Eiffel tower, for then the Eiffel tower, by the same token, would be about my thought. It does not do to say that the way in which smoke is about fire is a paradigm for the way in which my Eiffel tower thought it about the Eiffel tower, because fire always causes smoke (in the usual sense of fire and the usual circumstances), while the Eiffel tower only very rarely causes an Eiffel-tower thought in me. It does not do to say that my thought about the Eiffel tower is related to the Eiffel tower in somewhat the way in which the French locution 'la tour Eiffel' is related to the Eiffel tower, because the latter relation presupposes the first: only a creature capable of understanding the French locution, hence of having Eiffel-tower thoughts, can sustain the relation between the French words and the metal construction (a point sometimes made in terms of *derivative vs original, primitive, or intrinsic* intentionality).

In the context of CTM, the problem becomes that of understanding how the formal symbols in the system get to carry information about the Eiffel tower and other external entities, events or matters of fact. This is because for the system to have a thought about the Eiffel tower consists in the system 'activating' a (possibly complex) symbol which carries the appropriate information about the tower, and labeling it with the appropriate propositional attitude: belief, desire, regret, fear, etc.

There are actually two problems. One is to make intelligible the very idea of a representation in naturalistic terms. The other is to give an account of how a given internal (mental or brain) state gets to represent a particular thing (object, relation, event, fact,...). There is a functionalist attempt at solving the first problem, which goes roughly as follows. For the internal state S to represent X simply is to play the appropriate role in a causal nexus which terminates at the perceptual and motor interfaces. I think it is fair to say that this attempt has failed to elicit a broad consensus. More progress has been made on the second problem, yet even on that front there is hardly a consensus that we are nearing a solution. There is also little agreement on how important it is to solve the problem, on what would count as a solution, or even on whether it is well-posed.

There are, it would seem, at least four broad categories of mental representations. The first represent objects, the second represent n-ary relations ($n = 1, 2, \dots$), the

third represent logical constants, and the fourth represent states of affairs. The discussion is initially restricted to concrete entities: this particular object, the particular predicate ‘is a cherry’, or ‘is red’, the singular fact ‘this is a red cherry’. And it leaves to the side the issue of nonconceptual content, which concerns representations, if they exist, which are not propositional (more about this in the next subsection).

Surprisingly perhaps, what gets most attention in current discussions in philosophy of mind is the seemingly most complicated kind, *viz* representations of states of affairs. Several theories are at hand which purport to account for the relation between a mental representation and the state of affairs which it represents (for example: ‘There is a predator closing in on me’, or ‘This is a horse’); or to put it in causal or genetic terms, an account is sought of how a given material state or event gets to carry the information that there is a predator in front of me, or this is a horse. One strategy is to start from clear biological, psychological or engineering cases in which there is no reasonable doubt about a component, a state or a process *carrying* some information regarding the external world. The aim then is, on the one hand, to generalize from these unproblematic examples to an overarching informational theory of mental content or intentionality (or semanticity as one sometimes says); and on the other, to give a principled characterization of the relation which discharges a number of crucial obligations, the most central of which is to make intelligible the possibility of misrepresentation. This is the proper domain of informational (Dretske) and teleo- (Millikan, Papineau) semantics.³³ The intuition common to these theories is that some situations must be identified by a creature if it is to survive or at least to behave adaptively, and that those situations get to be represented mentally. The possession of a disposition to form a representation of one such situation is a biological trait of the creature which results from either ontogenetic learning or phylogenetic selection. Whatever the merits of these proposals, which cannot be discussed here, it is far from clear, at this point, whether they generalize beyond a set of very simple concrete, ecologically significant situations, which are indeed likely to be detected by evolved creatures.

But how do n-ary relations ($n = 1$ or 2 will do) get to be represented? Another way of asking the question is, How does a mind acquire a concept? One answer is, by learning from examples and counterexamples (this is what the field of machine learning is about). This answer again, whatever its merits, remains incomplete, for no learning is possible without the possession of more basic concepts. It is so difficult to imagine how basic concepts could be acquired that some authors have come to believe that they are not, that they are innate. But this makes the first problem of intentionality more pressing: having some idea of how a particular thing gets to be represented mentally may lead one to begin to grasp what the representing relation consists in; but in cases where there is *no* imaginable process by which a particular thing (in the case at hand, a concept such as ‘is a cherry’ or ‘is a solid, permanent, rigid object’, or a relation such as ‘is taller than’ or ‘is hidden behind’ or ‘is a part of’, or ‘is caused by’) *gets* to be represented mentally, then one has no initial grip on the problem of what it is for a mental state to represent the concept ‘is a cherry’ or ‘is caused by’.

Leaving untouched the issue of logical constants, and the problem of abstract and of non-existent entities, I conclude this section by returning to the matter of what a language of thought can represent. I said that, with an important proviso, it could conceivably represent anything at all. The proviso is: anything which is representable as a finite combination³⁴ of its primitive symbols. Seen semantically, this means that only those entities, relations or facts which are expressible within the bounds of the conceptual apparatus of the creature under consideration, can be represented by that creature. This seemingly rigid limitation is seen by some critics to run against the evidence of an unbounded capability of humans to create new concepts and generate new thoughts (a capability which precisely goes beyond the creativity which LoTH accepts, and on the basis of which it actually is defended).

Yet our discussion of intentionality is not complete. It is time to introduce an essential new concept.

2.5 *Tacit Knowledge, Subpersonal Processes, Non-Conscious Mentality*

I have up to this point left implicit the assumption, encouraged by the toy examples provided, that what is represented in the mind is the sort of entity which typically populates our conscious, or even more specifically our verbalizable thoughts, and that the representations themselves are in fact either conscious or potentially so. That this has been thought, in certain contexts, to be a reasonable assumption is illustrated by the first phase of artificial intelligence. And that it can be essentially invalidated by experience is perhaps illustrated by the failures of this phase of AI³⁵. But AI paid a dear price for ignoring the lessons of the prehistory of cognitive science. From Leibniz to William Hamilton, Helmholtz, Bain³⁶ or Lashley, there are numerous rediscoveries of the idea that our conscious mental states do not form a connected flow, whether in a causal, a rational or a temporal sense of connectedness, and that there must be intermediate or subjacent states without which no account of our conscious mental lives will be forthcoming. As Bain puts it: 'Outward expression, however close and consecutive, is still hop, skip and jump. It does not supply the full sequence of mental movements'.³⁷

In the rise of cognitive science, this insight played a decisive role and is generally credited to Noam Chomsky, who postulated that linguistic ability is based on the speaker's *tacit knowledge* of syntactic rules.³⁸ Daniel Dennett developed, in his influential first book,³⁹ a distinction between 'personal' and 'subpersonal' states and processes. Stephen Stich (1978; see also 1983) coined the expression 'subdoxastic' to characterize mental states which lead to genuine doxastic states such as full-fledged conscious beliefs. Douglas Hofstadter developed the notion of 'subcognition' in a paper dated 1982.⁴⁰ The rise of the connectionist approach, in the wake of a famous book subtitled *The Microstructure of Cognition*,⁴¹ led Paul Smolensky to draw the contours of a 'subsymbolic' level of description.⁴²

Throwing all of these notions into one basket seems to beg for trouble, as they are not just different presentations of the same idea. They do however have three crucial features in common. First, they are both related to and different from regular propositional attitudes as traditionally construed. Second, they are assumed to play an essential role in the genesis, attributes and dynamics of these regular mental states, as well as many if not all other aspects of our mental life. Third, they are proffered as genuine *cognitive* entities, and not as merely *physical* (neurobiological) entities.

As to the first point: These states or processes differ from regular propositional attitudes in one or more of the following dimensions. They are non-conscious; they may be inaccessible in principle to consciousness; they may defy description in ordinary language; they may be devoid of a logical structure. As to the second point: these states and processes are assumed to be either causally and/or explanatorily essential to the nature and regularities of our conscious mental life, particularly its more rational province populated with stateable beliefs, desires, fears, regrets and so forth. They are assumed to fill the gaps noted by Bain, but even more importantly they are thought to provide a *uniform* account of cases which seem to only require standard rational explanations *and* of cases which resist the treatment. The ‘micro-cognitive’ approach (using this neologism as shorthand for the various departures from ‘fully’ cognitive in the traditional sense) is also applicable beyond the realm of propositional attitudes, to perceptual or motor states for example. In fact, the discussion of the 1980s is continuing now in the context of mental states with *non-conceptual* (typically but not exclusively, perceptual) content.⁴³

The third and final feature of these entities is that they are not *simply* brain processes or neuroscientific constructs. There is always a difficulty in stating precisely what this is supposed to mean, for, as methodological naturalists, cognitive scientists, like naturalistically inclined philosophers of mind, do not doubt that in the last resort, mental processes of any sort *are* brain processes. The point is that, just as is assumed about *cognitive* states, there is an essential feature which microcognitive states possess and which is not captured in the extensional vocabulary of biology as it is now conceived. (This is *not* as mysterious as it might seem: a 100-euro bill is, beyond doubt, a piece of paper, but it is also endowed with a function which paper science does not capture). They are *also* endowed with a function which, although accomplished by strictly neurobiological means, is *cognitive* in an extended sense. What this amounts to is a contentious matter. Friends of cognitive science (as we know it today) appeal to a generalized sense of ‘intentionality’, which extends the usual relation to cases where the target or the content of the ‘vehicle’ (the entity inside the brain which is the first argument of the relation) can be something which does not belong to the usual ontology, *i.e.* is not picked by words, or concepts, or percepts to which we have unaided conscious access. They believe that microcognitive states can be *informational*, that they carry information which can be manipulated and combined so as to yield the desired conscious mental states or observable behavior, as LoTH postulated of cognitive states.

Skeptics such as Hubert Dreyfus or John Searle⁴⁴ see no hope of making sense of a general notion of information which would not appeal to a conscious human agent. In other words, they deny the reality, or the usefulness, of a level floating

mid-way between full-fledged mentality and brain events (or, perhaps, electronic or other physical events). These skeptics about information *ipso facto* harbor doubts about cognitive science as deployed today, insofar as it may be surmised that any attempt to account for full-fledged, personal-level, conscious mentality in terms of something else which is not simply brain facts, will meet with their disapproval. And it can be argued that cognitive science as we know it parts ways with all previous epochs of psychology, as well as with commonsense psychology, despite numerous historical and conceptual affinities, precisely by giving pride of place to this intermediate level which is still mental but not conscious, nor translatable into behavioral dispositions, nor (as in Freudianism) systematically correlated to states which could intelligibly be interpreted as conscious.

Both sides take the explananda of scientific psychology to include mental phenomena as they appear to the untutored eye, embedded in a 'folk theory' nourished by introspection, by familiarity with people's sayings and behaviors, and by first and third-person reports, whether purportedly real or fictitious. The disagreement concerns the language in we should expect the sought-after explanantia to be couched. The first camp conceives of it as essentially different from, the second camp as essentially commensurate with the language of folk psychology, which is intentional through and through. The eliminativists in the first camp believe that psychology must seek accounts of mental phenomena in the usual sense in terms of something different, on pains of being sterile, while the anti-eliminativists in the second camp see this strategy as self-defeating, providing at best explanations or descriptions of something else; in a nutshell, as changing the topic. One side wants to purge scientific psychology of any folk or naïve concepts, the other wants to preserve a semantic core of folk psychology while of course refining and completing it. Note that the issue of reductionism cuts across this divide: some eliminativists want to deny any sort of reality to folk theory, so that to them there is nothing to reduce to some more basic level; and many anti-eliminativists are physicalists, who postulate a full ontological reduction of the mental realm to basic physics.

2.6 *Consciousness*

Cognitive science, and philosophy of mind as we know it today, owe their existence, to a large extent, to the rejection of behaviorism's rejection of mental states. Yet cognitive science is also the heir to behaviorism, and one telltale 'hereditary' sign is that for a very long time it did not regard consciousness as part of its agenda. Philosophy of mind, by contrast, did, but the strategy, following Dennett's recommendation⁴⁵, was to tackle intentionality first, in the hope that consciousness would fall out of the solution to that problem. Consciousness did play a marginal role in some experimental paradigms, but as a phenomenon to be studied it was reduced to attention, sometimes referred to as the 'gateway' to consciousness.

There was a rather abrupt change in the years 1975–1980, when almost simultaneously consciousness appeared, on the one hand, as presenting some deeply puz-

zling, strongly counterintuitive features, and hence as requiring scientific attention, and on the other hand, as unsuited for scientific study. So just as evidence of a ‘cognitive unconscious’ was mounting, some spectacular manifestations of which are priming (unconscious perception), blindsight, or unconscious memory⁴⁶, it dawned on some philosophers that for broadly logical reasons, conscious experience, the first-person phenomenon *par excellence*, could no more be studied from the third-person viewpoint which is constitutive of science than teeth can bite themselves.⁴⁷ Ned Block proposed to distinguish two ‘kinds’ of consciousness, the first of which, ‘access consciousness’, he assumed to be open to third-person investigation, the second, ‘phenomenal consciousness’, containing (in both senses of the word) the seemingly irreducibly first-person component.⁴⁸ Whether this is a valid distinction or a valid strategy remains hotly disputed. David Chalmers defends the view that there are ‘easy’ and ‘hard’ problems of consciousness, the latter forcing one into a position which he calls ‘naturalistic dualism’.⁴⁹

There is now a quasi-autonomous field of ‘consciousness studies’ where the disciplines which are involved in cognitive science play a leading but not an exclusive role.⁵⁰ Two of the new players are quantum physics, and that part of philosophy of mind which has no privileged relation to cognitive science. This autonomy however seems to reflect historical, institutional and other social-psychological factors, rather than a natural division of labor. This is not to say that consciousness does not constitute a special area within cognitive science, somewhat in the way cosmology, although part of physics, is a special part of it. Nor should we assume that consciousness is for cognitive science a ‘normal’ problem which is bound to yield to systematic efforts; the arguments to the effect that it might not are not easily deflected⁵¹. But intentionality, as we saw, gives rise to a similar, if perhaps less radical doubt, and so does, according to Fodor, ‘belief fixation’ in his sense. In other words, cognitive science accommodates problems *and* mysteries, which is why philosophy makes such an important contribution.

On the other hand, there remains a tension between those who see consciousness as a problem to be solved (or dissolved) by cognitive science with resources commensurate with those which it deploys at present (including mainstream naturalistic philosophy of mind), and those who believe that the lack of progress in accounting for the subjective, phenomenal character of conscious experience presages a conceptual revolution, in the wake of which the entire edifice of cognitive science would be so thoroughly revamped as to become nearly unrecognizable.

3 Assessing the Framework in the Light of Cognitive Science and *Vice-Versa*

The entire discussion of Sect. 2 provides a partial characterization of a family of related views about the mind, centered around the CTM and the LoTH, but allowing possibly important departures from them. This family of views, or this basic approach, provide what I will refer to as the liberalized classical framework (LCF)

for cognitive science. There are two questions which must now be asked. One concerns the adequacy of the framework with respect to cognitive science as it ‘really’ is: How faithful, how informative, how useful? The second concerns the plausibility of the framework: Is it internally consistent, is it orienting the field in the right direction, *i.e.* does a true picture of the mind plausibly lay in the direction towards which it is pointing? The two questions –descriptive and prescriptive– are not independent. Cognitive science cannot be grasped as it ‘really’ is without the help of some principle of interpretation, which cannot be re-invented entirely anew for the purpose of assessing the framework under scrutiny: there is the familiar problem of the Archimedean point. As for the probability of being on the right track, it cannot be assessed independently of an evaluation of the results and failures of cognitive science as it is today: the proof of the pie is in part in the eating. To illustrate the predicament: An essential feature of the framework is that the mind is to be understood as an informational-computational system. How can we evaluate this proposal against what is being actually shown in cognitive science, when the criterion of a successful piece of research in the field is that it throws light on some mental phenomenon by revealing its informational-computational nature, where ‘information’ and ‘computation’ are to be understood in the way fixed by the conceptual framework?

There are however ways of breaking these two entangled circles, by some familiar back and forth maneuvers. Cognitive science provides explicit and implicit arguments pro and contra the framework, which in turn is needed to give meaning to the concepts which these arguments employ; the framework is thus thrown slightly off balance, yielding a slightly different view of cognitive science, and so on. The second circle is more problematic, as illustrated by the fate of previous epochs in psychology, such as behaviorism, which flourished until it all but vanished (or so it seemed) from the academic scene. However the issue raised is abundantly discussed in the debate opened by Kuhn’s 1962 *Structure of Scientific Revolutions*, and this is not the place to explore it. Perhaps the history of cognitive science, as seen by future generations, will provide an interesting new case study.⁵²

3.1 *The Incompleteness of LCF*

Perhaps the single most important fact regarding the status of cognitive science is that it is not simply the unfolding of LCF. First there are cognitive scientists who explicitly reject LCF. But second, and more importantly, despite its being quite substantive, and thus open in principle and also in fact to destructive attacks, LCF leaves a considerable amount of freedom to anyone wishing to develop an empirical science of the mind. Although more than a ghost of a theory, LCF provides at best a skeleton, and there are indefinitely many ways of building the full creature around it.

A simple way of making apparent the abyss between LCF and a worked-out research program for cognitive science is to ask the following schematic question: How does one go about finding out about X, where X = memory, learning, linguistic

and communicative skills, musical ability, concept acquisition, language learning, reasoning, problem-solving, social competence, navigation, vision, audition, motor control, face recognition, special skills such as driving, playing chess or baseball, etc., and where ‘finding out’ can mean providing an explanatory account, a model, a functional analysis, a evolutionary rationale, a neural basis, a clinical manual? LCF provides precious little guidance.

LCF does provide an important constraint on what counts as a valid contribution to any of these undertakings. In particular, what counts as a ‘model’ is under pervasive influence of LCF. A model can mean a computational model in the strict sense, *i.e.* a piece of software or a dedicated digital device or an (artificial) neural net; but in general ‘model’ can take on an indefinite variety of other meanings, and the LCF is *prima facie* all-important in placing limits on this variety. Acceptable models meet two broad specifications: they treat the ability to be modeled as a particular way of processing *information*, and this processing must be shown either directly or indirectly to be mechanizable by well-identified or at least conceivable means. Indeed, the first label of what can be retrospectively seen as the beginnings of cognitive psychology was ‘information-processing psychology’. However, the sense of mechanism is quite broad, leaving a lot of elbow-room to the modeller. First, there is no requirement that some machine be proposed which would realize or implement the X at hand; it suffices to show that the performance of a case of X can be broken down into sub-tasks or processes Y_1, \dots, Y_n , each of which is shown, or assumed to be, itself mechanizable in the same sense (with a general clause forbidding a non-ending decomposition). Second, perhaps paradoxically, while the strict computational framework demands that the mechanizable processes be computations in the technical, Turing sense, the models accepted by LCF are just called computational without their proponents necessarily having any clear idea of what a computation is in the technical sense, or finding it at all relevant. To them, computational just means mechanizable in principle in the foregoing sense. An important case in point is the neuroscientific or brain models of cognitive functions: they are informational alright, but they are hardly ever computational in any sense recognizable by a logician or a computer scientist. To some neuroscientists, the further obligation of showing that the brain mechanisms involved in the model are computational in some well-defined sense⁵³ is important and must eventually be discharged by theoretical or computational neuroscientists, on pain of rejection of the proposed model. But others don’t think this is necessarily called for, and would be loth to reject a model on the sole basis of a failure, on the part of computational neuroscientists, to come up with a plausible computational account.

To sum up, the constraints which LCF imposes on what counts as an acceptable model raise issues of interpretation. These constraints remain quite important nonetheless, as they fix a basic level of theoretical description (the informational level) while maintaining a demand for physically realizable mechanisms. But beyond that, LCF provides not the slightest cue when it comes to attacking any particular mental function, ability or process. The job of the philosopher of cognitive science cannot stop at discussing LCF and its limitations: it must examine the ways in which cognitive scientists, and philosophers of mind working with or among them, seek actual

scientific results. This does not imply an exclusive focus on the most local issues (say, non-nutritional sucking *vs* gaze duration as a method of eliciting neonates' perceptual and cognitive abilities, or prototype *vs* exemplar theories of concepts, or ventral *vs* dorsal visual pathways, or phenomenal *vs* neurophysiological accounts of pain, or weak central control *vs* metarepresentational theories of autism); all scales and levels of generality cry out for examination. The point is that the highest level, where LCF is discussed, is by no means the only one, and as far as descriptive philosophy of cognitive science goes, may not even be the most important one.

In the remainder of this chap. I will try and compress some examples of theoretical and methodological issues at various scales which contribute both to operationalize LCF and to assess its role and legitimacy as a foundation for the field.

3.2 *Theoretical Issues. Some Examples: Modularity, Innateness, Reasoning and Rationality*

The single most important and controversial theoretical issue at a level of generality sitting close to, yet below LCF bears on the 'architecture' of the mind. As we saw, one argument in favor of CTM is that there exists a universal Turing machine, an abstract device which can be regarded *post facto* as an idealized model of a programmable computer. One such machine can simulate any other Turing machine, if it is provided with its table (in computer jargon: a programmable computer can compute the results which any given dedicated computer grinds out, once the wiring pattern of the latter is translated into a program fed to the former). Similarly, the universal Turing machine might be regarded as a model for the mind as an information-processing system, which would account for the apparent ability of the mind to perform any imaginable operation, regardless of subject-matter; in other words, the mind, on that view, would be comparable in some respect to a 'general-purpose' computer. This is but the first ingredient in the elaboration of a proposed architecture of the mind which early AI sought to substantiate, and which is also akin to folk-theoretical or traditional views. The mind, under this proposal, is completely impartial as to the tasks which befalls it: whatever learning, memorizing, retrieving, inferring, comparing, etc. are required to perform a task, whether visual, arithmetic, motor, linguistic etc., whether it involves people, cars, animals, abstract entities, etc., they are just the general processes of learning, memorizing, retrieving... which the mind is equipped with and which it deploys as the need arises. In simplistic terms, the mind stands to cognitive tasks somewhat like a pick-up truck stands to moving medium-sized objects: it gets them loaded in one place and carries them in another, and neither the nature of the objects nor the points of arrival and departure make any theoretical difference.⁵⁴

It has been argued that this architectural proposal is deeply mistaken. Chomsky has defended a notion of a 'language faculty' clearly individuated within the cognitive system, somewhat like a bladder or an arm are clearly delineated organs of the body. He has argued, more generally, leaning on both logical and empirical consid-

erations, against Piaget and Putnam among others,⁵⁵ that no purpose-general learning procedure can lead a baby to acquire her mother tongue, and has suggested more generally that the very idea of purpose-general learning is a flawed concept, or one with limited use in the study of basic cognitive abilities. Fodor, in an epoch-making though slim book,⁵⁶ had proposed to distinguish two broad kinds of cognitive processes. The first, which he called 'input systems', are akin to Franz Gall's faculties, and are accessible to scientific study. The second are 'central processes' which may well be forever hidden from the scientific eye. The distinguishing trait of input systems is their 'modularity'. Fodor drew a list of features which tend to cluster and constitute partial criteria of modularity. The discussion of this list and of various possible understandings of what a module truly is (in particular, how important to the proposal is the possibility, or the demand that a module be *physically* locatable in the brain) has generated an immense literature, which continues to this day.

Characteristically, 'lower' faculties which we share with other animals are modular: the various perceptual modalities and motor control are modular; in addition, linguistic competence (or at least certain dimensions or stages of it) are modular. By contrast, 'higher' faculties which Fodor groups under the heading 'belief fixation' are non-modular. It is worth noting that this division corresponds to a large extent to the personal/subpersonal distinction discussed above (§ 2.5), and also to the boundaries of the folk-psychological domain: as much as we seem to have an untutored grip on 'belief fixation' (how we come to believe what we do believe, in the usual explicit, verbalizable, introspectible sense), we have no clue on how the visual system delivers, on the basis of the retinal array, intelligible scenes such as an unknown human face perched on top of a human bust reclining in an armchair, or the bus arriving at the stop; or on how, from a string of air vibrations, we instantly infer that we are asked to pass the salt, and so on. Thus Fodor's position in the two-camp situation described at the end of §2.5 is this: psychology is free to try both the subpersonal reductive and the personal non-reductive approaches, but only the first seems to work, and then only for topics where there is no personal-level processual phenomenology, just personal-level end products. (In Bain's vocabulary, there is just one big 'hop, skip and jump' between the 'pure' stimuli, inaccessible from the first-person viewpoint, and the conscious perception or understanding).

This pessimistic outlook provoked a reaction from some thinkers engaged in the study of higher processes, and they developed arguments in favor of 'massive modularity'.⁵⁷ They conjecture that modularity is a massive characteristic of the human cognitive system, not one restricted to peripheral and other automatic processes. Their defense generally combines evidence from developmental studies, showing that very small children, and indeed other animal species, exhibit high capabilities in certain specific domains clearly related to higher processes in adults and clearly unlearned in any normal sense of the word; neuropsychological evidence, showing that certain higher processes are regularly selectively impaired, psychological studies of patterns of performance in normal adults, which tend to show 'informational encapsulation' (segregation of bodies of information, preventing a module from exploiting knowledge not belonging to its proprietary 'data base'); evolutionary considerations; anthropological studies of deep invariants and ethological evidence

and speculations about basic cognitive needs of the emerging *Homo sapiens*; and finally conceptual considerations from complexity (the need for the system to overcome the combinatorial explosion).

What emerges as the most important characteristic of massive modularity is ‘domain-specificity’, the conjecture that the human mind applies distinct sub-systems to handle different regions of its environment (different ‘task domains’), from one–one interactions with other people and social behavior to handling middle-sized objects, categorizing living beings, or counting. This raises the following difficult question: Is the mind entirely modular? If not, then aren’t we left, for the non-modular part, with essentially the same problems which have beset the study of higher functions since the beginning of psychology and led Fodor to his pessimistic conclusions? And if the mind is entirely modular, how can one account for its flexibility and its ability to bring together vastly different realms?⁵⁸

The discussion of the central thesis and these related issues is inconclusive at this point,⁵⁹ but it is a goldmine of questions which are of independent interest. This is an example of an area in which the shoulder-to-shoulder collaboration between philosophers and empirical scientists is most clearly productive.

Massive modularity clearly implicates the issue of nativism, which has a much longer tradition of its own. The question has undergone a quiet revolution as a result of recent work in evolutionary theory, from biologists as well as philosophers of science, advances in developmental psychology, evidence from anthropology and linguistics, progress in neuroscience and in genetics. Perhaps the single most striking fact to emerge from this body of work is that, on the one hand, any simple notion of innateness which might be proposed is inadequate, and that instead several distinct notions may be required—a fact which critics will see as counting against the entire nativist undertaking—, while particular research programs broadly consonant with nativism are thriving.⁶⁰ Naturally, nativism is not thereby vindicated, and indeed it runs into harsh attacks,⁶¹ in particular from the ‘constructivist’ camp,⁶² but even then the conclusion that there is little worth retaining from the research would be unwarranted: controversy is the norm in science, rarely the sign of a mortal crisis. Unfortunately, this point, though obvious from the history, and present state of mature sciences, is not always understood when it comes to cognitive science, the result being an exaggerated suspicion towards the field.⁶³

One of the least likely faculties to succumb to domain specificity would seem to be reasoning: the very idea of reasoning seems to imply domain-generalty. As a final example of a theoretical issue which concerns to an equal degree psychologists and other cognitive scientists and philosophers, I will say a few words about the psychology of reasoning and the rationality debate. Decades of experimental research have shown that people fail at certain simple reasoning tasks, both deductive and involving uncertainty, despite being put in ideal conditions. Further, insofar as these failures affect the ability to reach good decisions, in real-life conditions, they are sometimes thought to establish our ‘irrationality’: we are unable to optimally adjust our means to our ends. Finally, it has been more recently shown that by manipulating the context, or cognitive setting, in which the tasks are performed, so as to make them less general and world-detached than the supposedly simpler,

leaner original setting, then the performance can rise dramatically. (This summary ignores the considerable differences between deductive and ‘inductive’ or uncertain reasoning).

A bird’s-eye view of the field reveals three major issues of immediate interest to us. Closest to the topic of the present section is domain-specificity. It has been proposed that humans are equipped with a special-purpose ‘cheater-detecting’ mechanism which allows them to perform quite well on certain deductive tasks, provided the context of the task makes salient a particular cheating condition.⁶⁴ This mechanism would have evolved in order to enable stable cooperative arrangements which gave our species a selective advantage in the ‘EEA’ (environment of evolutionary adaptedness). If this highly speculative, and much disputed hypothesis were confirmed and if other broadly consonant hypotheses, covering other aspects of reasoning, were formulated and supported, one would be in a position to argue that a domain-general ability such as (abstract) reasoning is parasitic on a domain-specific, evolved competence; that the latter is fast and reliable, while the former is slow and prone to failure; and that in some situations a transfer mechanism can allow the fast, reliable, specialized mechanism to trump the slow, mistake-prone general method and procure the needed solution in an area different from the one for which it evolved.

This conjecture is related to two more general ones which have received broader attention. The first is that contrary to what logicians and rationalist philosophers have taught us, reasoning is not based on formal logic, but rather on content-processing manipulations. ‘Reasoning without logic’ is the slogan brandished by Philip Johnson-Laird, arguably the most influential psychologist of reasoning and the creator of the theory of ‘mental models’.⁶⁵ A critical minority⁶⁶ argues that this is neither shown by the experimental results nor conceptually coherent. However, if we cross the domain-general/domain-specific opposition with the formal / content-driven (or syntactic/semantic) contrast, we get a two-by-two logical space in which the mental models theory (and related proposals) occupy an interesting position, combining domain-general with content-driven. Now LoTH being a syntactic theory of mind, it would appear incompatible with Johnson-Laird’s views, but this is hardly the case: Johnson-Laird has written a forceful textbook devoted entirely to a view of psychology based on the assumption that LoTH is true.⁶⁷ Similarly, many defenders of massive modularity accept LoTH, which one might not have expected as LoTH sits so well with domain-generality. This only serves to illustrate the point made earlier regarding the intended meaning of LoTH, and the problems cognitive scientists run into when they try to bring it into consonance with their own practice. This is true a fortiori of LCF, which leaves unspecified some degrees of freedom with respect to LoTH. Cognitive scientists are thus justified in leaving it in a semi-interpreted state, expecting its full meaning, and its true worth, to emerge from future developments in cognitive science.

The second important idea to which the cheater-detection hypothesis is linked is that of real-world, as opposed to ideal rationality. While the latter consists in making optimal decisions in all circumstances, including the most difficult from the theoretical standpoint, the former is geared to the set of ecologically likely situ-

ations, and produces in probability the best compromise between feasibility and efficacy. The feasibility condition was first proposed by Herbert Simon under the label ‘bounded rationality’,⁶⁸ where it went hand-in-hand with domain-general⁶⁹. In the framework of a modular architecture of the mind, an assemblage of domain-specific, evolved capacities, efficacy is relativized to particular sets of events or choices which are likely in the creature’s ecological niche, and may well not include theoretically worst-case scenarios. It has been shown⁷⁰ that very simple (‘fast and frugal’, or again ‘quick and dirty’) heuristics can do as well or better than elaborate and cognition-intensive reasoning strategies in certain families of real-world situations. Whether these heuristics are less brittle than the AI-inspired expert systems which made the headlines in the 1980s remains to be shown, but the general principle of taking into account ecological priorities in the overall rating of a given notion of instrumental rationality is worth exploring. It is connected with two yet wider perspectives. One, defended most vigorously by Stephen Stich,⁷¹ concerns philosophy and rejects the traditional method of conceptual analysis as a valid way of fixing the meaning or reference of such terms as ‘truth’ or ‘rationality’: empirical work is needed to discover what humans, across the globe, take truth or rationality to be, and cognitive-scientific methods, including evolutionary considerations, can help us interpret the results to yield a naturalistic description of such ‘thick’ normative terms. The second concerns psychology: the rationality debate has focused on the ‘bleak implications’ interpretation of the empirical data regarding systematic failure on the part of subjects to respond to problems in inductive reasoning (especially, although deductive aberrations are also implicated) according to the norms set by probability calculus and logic. The bleak implications view holds human rational aptitudes to be severely deficient. It contributes to a generally harsh assessment of human capacities, also based in part on a negative evaluation of the role of emotions. The purpose of ‘positive psychology’ is to discredit the general aim of finding faults in human cognitive performance.⁷² Humans should be judged, according to this new orientation, not on their performance in laboratory and pen-and-pencil tests, but on their ability to cope with situations and problems arising in the real world, in their daily lives; and this performance again is not to be assessed against the standards of abstract academic thinking, but by reference to success and failure, as perceived by the agents and their peers. Cognitive science is thus invited to step down from its normative pedestal and take a more descriptive stance, somewhat in the way in which philosophy of science, one generation back, took its historical-descriptive turn.

3.3 Methodological Issues: Paradigms, Levels and Reduction

The third broad family of issues which keeps philosophers of cognitive science and at times cognitive scientists busy concerns the comparative assessment of competing and seemingly incompatible conceptions of what counts as a theory, a model or an explanation in cognitive science. These issues can be framed, and often are

framed, as ontological questions, and are discussed alongside such central issues in philosophy of mind as representations, information, intentionality, consciousness, externalism, with which in fact they overlap. From that perspective, one asks, for example, whether the mind really is functionally equivalent to a programmable computer, or rather to a neural net, or again to a dynamical system; or whether the brain really is a biological computer, or some other kind of system. However it seems more profitable to view such questions as methodological. There is indeed no such thing as a biological computer, or an information-processing system, or for that matter a system, whether complex, dynamical or whatnot, independently of a theoretically driven description. Thus one might as well go straight for questions which are typically meaningful for philosophy of science: How do the various available proposals for describing the mind, for explaining and predicting its dynamics, compare? What are their fundamental tenets, what are their goals, to what extent are they mutually exclusive?

The main contenders in the last 20 years have been the so-called symbolic and connectionist ‘paradigms’. The symbolic, also sometimes called the ‘classical’ paradigm (briefly: ‘classicism’), is best represented by LoTH. The connectionist paradigm is defined by reference to neural nets, to which it stands in roughly the same relation as the Turing machine stands to classicism. The most detailed exposition of connectionism, which comes in many versions, is given in the two-volume collective work entitled *Parallel Distributed Processing: The Microstructure of Cognition*, already mentioned, and more particularly in the seminal paper ‘On the Proper Treatment of Connectionism’, and subsequent writings of Paul Smolensky;⁷³ I will use PTC, after the title of his paper, to refer to Smolensky’s formulation and use it as a ‘representative’ of the connectionist paradigm on par with LoTH on the other side.⁷⁴ Both are complex mixes of descriptive and prescriptive views of what cognitive science is or should be doing, both are clusters of doctrinal theses concerning the true nature of cognition, its typical operations, the ‘architecture’ of the mind, the proper basic level of explanation, and how it relates to other possibly relevant levels. Both can be extended by making additional choices, such as between ‘empiricism’, a view often thought to sit better with connectionism, and ‘rationalism’ (here a form of nativism), better accommodated, it is claimed, by classicism.

PTC is in agreement with LoTH in conceiving of cognition as a material phenomenon best characterized as information processing, where processing is a kind of mechanical computation transforming content-bearing material entities. But whereas LoTH is committed to inference rules operating sequentially on language-like data structures, PTC locates the causal transactions at a level where discrete operations are but limiting cases of continuous operations consisting in registering and modulating activation levels. The system, prompted by some input, runs these operations in a massively parallel fashion, with no central command; it reaches a temporary equilibrium, which is interpreted as the result of the system’s processing of the input. In an important family of cases, the entire transaction can be understood as the application of a negotiating procedure, extended over time, between competing ‘soft’ rules (rules which can be violated under the pressure of contrary

demands). Or to use a conceptual scheme drawn from thermodynamics, the system ‘seeks’ a final state which maximizes the overall ‘harmony’ between the built-in soft rules embodying the system’s knowledge, given the initial exogenous impulse.

The architecture of a PTC cognitive system is modeled after a network of so-called formal neurons, which are akin to the simple threshold automata introduced by McCulloch and Pitts in a paper which launched the connectionist movement in 1943.⁷⁵ Devoid of central coordination, and operating in parallel, such networks come in many kinds of configurations, with sometimes important differences between them, but they all differ more radically from Turing machines or their real-world counterparts von Neumann machines (programmable digital computers as we know them).

The third contrast concerns representations, in both their syntactic and semantic nature. First, only LoTH actually uses syntax, in the sense of formal systems, which coincides with the appropriate level of causal analysis of the system’s dynamics, while PTC retains only the notion that causal regularities (‘form’), without syntax, can be analytically distinguished from ‘content’. In PTC, the form is a combination of architectural constraints and numerical parameters, either semi-permanent or transient. Content is distributed over semi-permanent parameters (‘weights’, playing the role of synaptic weights) and transient parameters (‘activations’). In LoTH, content is typically expressible in linguistic terms, although as we have seen LoTH can accommodate ‘subdoxastic’ content which corresponds to no ordinary concepts or words. In PTC, content is typically ‘sub-symbolic’ or ‘subpersonal’, and ‘personal’ content, *i.e.* thinking material used in everyday speech and conscious trains of thought, emerges at the global level of functioning of the entire system.

Fourth, for PTC the typical cognitive process is a form of literal or generalized perception, the identification of a previously registered pattern or the classification of a new pattern as akin to a stored one (in particular, a perceived pattern can be understood as a partial view and the processing consist in its completion). For LoTH, the typical cognitive process is a formal inference from complex symbols which are truth-evaluable (the ‘sentences’ of the language of thought). Thus, PTC seems to support a broadly associative psychology and LoTH a strictly inferential-logical psychology.

Related to the last, one more major difference deserves to be mentioned here. It concerns the place of learning. Neural nets naturally learn from exposure to experience, and store the knowledge thus acquired in the semi-permanent parameters (weights) regulating the mutual influence of the nodes upon one another. This learning process is an integral part of PTC, as shown by the fact that one could not even conceive of a neural net accomplishing some given cognitive task on the basis of ‘instructions’ received from outside: it has, or so it seems, to learn ‘for itself’. This is in stark contrast with LoTH, which countenances models whose cognitive functionality is due to ‘knowledge’ directly provided by the cognitive scientist, programmer or expert. To be sure, a classical system can be equipped with a learning device (the field of machine learning was mentioned earlier), but the point is that a classical system which has been directly fed the necessary knowledge is indistinguishable

from the same system which owes its knowledge to its learning component. A classical system can be understood quite independently from the process by which it has acquired its competence; in fact, the possibility that it might not have literally acquired it, except in the indirect sense of the distal causation of evolution, can be made sense of.

It would thus appear that LoTH and PTC advocate such widely different views of cognition and of what counts as a valid piece of research in cognitive science that at most one can be on the right track.

This has in fact been the understanding of scores of authors, who have defended their preferred approach by attempting to find a fatal flaw in the other, hoping thereby to win by default. Thus friends of PTC (and other brands of neural-net approaches) have found a whole series of lethal faults in LoTH, ranging from its rigidity and brittleness, its separate treatment of 'normal' and 'exceptional' inputs, its inability to naturally deal with context-sensitivity, to biological (both neuro-physiological and neuro-pathological) implausibility. Friends of LoTH have focused on problems affecting connectionist models such as scalability, applicability outside carefully crafted situations, to their purported inability to account for the essentially symbolic, systematic and generative character of thought.⁷⁶

However difficult and often technical these debates have been, they pale in comparison with a more basic issue: Is it the case that (at least) one or the other approach must be wrong? In other words, are they truly competing or are they rather partial views, from different standpoints, at different levels of description, and even possibly of different regions, of one large and complex realm of phenomena, so that they could both conceivably be at least partly right? The second alternative has by now almost become the received view, although for a wide variety of reasons.

It has been asked, first, whether connectionist models are not simply re-descriptions, at a lower level of aggregation, of classical models. After all, a computer running a program with a proven cognitive functionality can be described at a level where symbols, central control, sequentiality all disappear. Maybe connectionist nets are mere 'implementations', material realizations seen at an intermediate level of description, of classical models. This view is sometimes thought to be buttressed by formal results to the effect that neural nets can approximate Turing machines and that conversely, any net can, and in fact usually is simulated on a von Neumann computer. These implementational or reductive interpretations of PTC are ruled out, I believe, by the fact that contents cannot be re-described in a way compatible with the re-description of the computation processes, but this would require a long and careful examination.⁷⁷ In a similar vein, Smolensky⁷⁸ has proposed a well worked-out emergentist view in which, roughly speaking, classical models are descriptions in the limit of emerging behaviors of connectionist models. The levels issue is far from settled (and more will be said about it shortly).

Second, for many cognitive scientists, connectionism and classicism are simply providers of heuristics for constructing models, and models are not true or false, but better or worse in the sense of providing a more or less perspicuous mechanical account of the cognitive phenomenon at hand. Many believe that as a general rule,

hybrid models are the proper way to go: most cognitive functions, in their view, are carried out by complex assemblages of connectionist and classical subsystems.

But this in turn raises a general issue: How relevant are models for cognitive science? Of course, given the nearly vacuous content of the concept 'model' in contemporary scientific parlance, the question thus posed makes little sense. So let us return for a moment to a bygone era, that of artificial intelligence (AI), in the years from 1956 (its official birth date) to the late 1970s. AI then had a dual goal. It wanted to produce a general theory of human cognition, on the one hand, and on the other a model for each and every cognitive function, taken one by one. These goals were interlocked: the general theory inspired the construction of the specific models, and the models confirmed, and helped amend, the general theory. This proposal, reasonable as it may perhaps sound, nevertheless took on a highly unusual character due to the concept of model to which it appealed: a model of a cognitive function (playing checkers, or recognizing edges in a visual array, or proving the Pythagorean theorem, or translating from German to Russian...) was construed as a program, in the computer sense of the word, which allowed a machine to roughly equal the human performance on the given function or, in AI jargon, 'task'. Thus was born, and philosophically defended, the idea of 'programs as theories'.⁷⁹ Herbert Simon thought of AI and cognitive psychology as essentially identical, and another pioneer, Roger Shank, was fond of saying that AI was nothing but the study of human intelligence.

This view was quickly abandoned, though, for two reasons: first, in most cases, the programs proposed by AI were unintelligible (or provided no additional intelligibility beyond what had gone into their making in the first place); second, they did not perform well. A distinction was thus drawn between information-theoretic models for psychology or linguistics, on the one hand, and AI models for advanced computing, on the other. The general theory was elaborated (by philosophers) to become LoTH, and AI all but relinquished its early ambition of providing the 'glue' between machine simulation and psychological explanation: it chose to become an engineering discipline, period. And so cognitive science was left with a blander notion of a model: a psychological explanation of how elementary mechanisms, known or strongly conjectured to be a part of the basic equipment of the mind, can be strung together so as to produce a given cognitive function. Interestingly, this did not amount to an outright rejection of AI's initial proposal, which instead was trivialized. Mainstream cognitive psychology subscribes at least to LCF (if not to more stringent versions of functionalism), and thus it agrees with AI about the general nature of the explanation sought, *viz.* a string of computations on content-bearing (informational) units. But while AI was betting that the essential part of the explanation lay in the complexity of the program (the structure of the particular computation), cognitive psychological models hardly refer to that structure, which is usually about as rich and surprising as a bunch of connected watering hoses in an average backyard. The bulk of the explanation lies in the identification of the elementary mechanisms, and the precise delineation of the explanandum; computation is not mentioned at all, except sometimes as a label appended to the model, amounting to no more than 'mechanistically acceptable in view of our present state

of knowledge'; finally, *no* empirical support is drawn from the examination of computer simulations.

Connectionism has undergone a differentiation process partly reminiscent of the split between engineering AI and cognitive psychology. The engineering branch, now fully integrated into AI, uses artificial neural nets to simulate a variety of cognitive functions. The psychological branch, best exemplified by PTC, holds on to a strong notion of model which purports to offer both a standard of causal explanation in psychology (a general theory of cognition) and a method for constructing models of specific functions, just like classical AI, and attempts to resist trivialization by deploying, as we saw, a complex strategy resting on the postulation of a distinct level of description, lying between the 'wetware' and the symbolic level, that populated by typically introspectible and/or linguistically expressible thoughts and transitions. Finally, there is a 'neuroscientific' branch of connectionism, which sees itself as providing models of neural states and processes in the style of mathematical physics; in other words, connectionism in this sense is a style of modeling within theoretical neuroscience.

Whether these enterprises are essentially identical, complementary or largely independent is open to debate. However, the lesson which can be drawn from the earlier phase of cognitive science, where classical AI and LoTH held sway, is that a large part of the actual work on the various cognitive processes and faculties, at whatever level, is unaffected by the outcome: engineers will spend most of their energy on statistical analyses of data and of network performance and learning algorithms; psychologists will work on subjects' performance, employing the tools of empirical psychology, developmental psychology and neuropsychology; and theoretical neuroscientists will be looking at the fine details of certain neural processes and applying advanced modeling methods from statistical physics.

However, cognitive science has entered a new phase in the last decade, which can be seen as the third, coming after the pioneering phase dominated by the classical paradigm and the theoretical role of AI (mid-1950s to late 1970s), and the developing phase where psychology, allied to philosophy, played the leading role, while connectionism competed with AI as the dominant modeling approach. In much the same way as AI during Phase I, cognitive neuroscience in Phase III claims the central role, offering as main source of empirical evidence its proprietary method of functional brain imagery. The 'programs as theories' view of a cognitive-scientific model has been replaced by a 'dynamic map of the brain' conception, whereby once the areas involved in the production of a given (overt or covert) behavior have been identified, together with the flow chart of their activations, one has reached the desired understanding. Thus stated, the view raises several objections. First, there is an 'explanatory gap' between a brain map, whether static or dynamic, and the cognitive process or function which is concomitant with the imaged brain activity, just as there was an explanatory gap between an AI program and the mental function it was simulating. The gap can only be bridged by psychology and same-level disciplines. Second, the brain mapping comes in once the basic structure of the explanandum has been uncovered by psychological, linguistic, or other empirical work, buttressed by conceptual analysis, just as an AI program could be written (or should have been

written) only after psychology and others has cleared the ground. In short, functional imagery is neither the beginning nor the end of the process of discovery.

This should all seem straightforward,⁸⁰ and most cognitive neuroscientists are at pains to present functional imagery as a 'tool', albeit one as crucial to cognitive science as the telescope is to astronomy. Cognitive neuroscience in fact is often defined as the blending of psychology and functional neuroscience. But this leaves room for a range of reductionist positions which are in fact incompatible with the view that psychology and other disciplines concerned by cognition can continue to develop according to their own criteria and dynamics. There are three main sources feeding the neuro-reductionist current. The first is eliminativism, the belief that psychological concepts arise from our untutored intuitions and that they should be discarded wholesale in favor of a respectable conceptual structure drawn entirely from natural science.⁸¹ The second is fundamentalism, the belief that there is a unique ultimate level of description which captures the causal order and is thus the only literal truth-bearer, and to which neurobiology is closer than psychology. And the third is mechanicism, understood as a conception of the aim of inquiry (at least in the 'special sciences') as to provide mechanical models (in the contemporary sense of mechanism). There are a moderate and a radical form of neuro-reductionism. The moderate form recognizes as basic for cognitive science an integrative or emerging level which is relatively insulated from the cellular, molecular and lower levels, just as classical functionalism sees the informational level as relatively insulated from the physical level. The radical form⁸² sees no reason to set a principled limit to the search for ever more basic explanations, and makes the (controversial) case that cellular–molecular models are already at hand for some basic cognitive functions, showing the way to a reduction of the entire domain of cognition.

Whatever the merits of these views as visions of tomorrow's cognitive science, as interpretations of the current state of the field they run against two uncontroversial facts. The first is that there are few, if any, elucidations of a cognitive phenomenon which can be stated as a self-contained neuroscientific theory or 'model'. The second is that there are many important phenomena which we simply would not know how to attack as neuroscientific problems. This does not tell against the importance of neuroscience for cognitive science, or of imagery as an essential new resource; what is at issue here is the idea that neuroscience is wedded to *the* natural basic level of cognition and that it is therefore fated to produce a complete, self-standing theory of the mind. And this idea in turn is connected to the dubious assumption that providing a model is all we need to explain a given cognitive process or function.

Much depends, of course, on what cognitive neuroscience is meant to encompass. It has traditionally been thought of as combining neuroanatomical, neurophysiological, neuropsychological dimensions, as well as the relevant constraints from cellular and molecular neuroscience. Should it also include the relevant part of evolutionary theory? The considerable revival and extension of evolutionary thinking in cognitive science seems to show that, regarding the brain also, the Mayrian duality of proximal and distal causes must be countenanced. But most of the recent work in evolutionary psychology and allied disciplines has little to do with brain

science, although of course there are systematic attempts (not all theoretically motivated) to connect the two areas.

The unavoidable conclusion, it would seem, is that cognitive science is for now, and possibly forever, wedded to a form of explanatory and causal pluralism.

3.4 *How Can the LCF Survive in a Hostile World?*

Pluralism is a peace-making device, but it comes at a risk. First, of course, the unity of the field is under threat. Perhaps in an age where ‘disunity’ is seen by many as a fact of scientific life, this should not worry us. But second, the coherence of the entire enterprise is under suspicion: for what grounds are there for hoping that cognitive science will keep holding together conceptually disciplines and approaches which have a long tradition of separatism, if it is forced to renounce its monistic principles?

Cognitive science up until now has overcome these dangers. This it has achieved by containing the pluralistic pull within a monistic framework, while at the same time developing a federalist practice and ethos. The framework was initially provided by the CTM, which was then gradually liberalized to LCF so as to make room, on the one hand, for fairly precise alternatives to classicism (as explicated in LoTH) such as PTC and other varieties of connectionism, and on the other hand for the countless research programs sharing a family resemblance with one another and with CTM, yet having no clear theoretical connection with it but only a shared understanding of the nature of the problems at hand and of what counts as progress in their resolution. In fact, this is how I propose to characterize, functionally so to speak, what I have been referring to as LCF. LCF views cognition as based on, if not necessarily strictly identical with a set of functionalities of a complex biological system (the central nervous system) within the human organism and by extension within other organisms. Cognitive science is thus seen as a branch of biology, giving rise to a non-threatening structure/function dichotomy, more recently doubled by a distal/proximal duality which is the hallmark of evolutionary thinking. LCF proposes to express regularities at the functional level in terms of information, where information is a place-holder (much like what ‘force’ was for a long time in physics, or ‘gene’ in biology) to be gradually filled as cognitive science develops. Whatever scientific concept will eventually be regarded as meeting the specifications, the immense resources of commonsense disciplined by philosophy, logic, linguistics, economics and other sciences of man can meanwhile be tapped to provide partial descriptions and explanatory schemas regarding cognitive dynamics, for they all variously circumscribe areas of informational transactions.

LTC thus affords both a unifying framework and the leeway required by the many enterprises which tend to one or the other of the countless manifestations of cognitive activity without committing themselves to any particular theoretical option regarding cognition *in toto*. An imperfect historical parallel is provided by the physical sciences in the 18th century, which all accepted the notional umbrella

of Newtonianism, yet felt free to develop their proprietary ontologies and practices without feeling the need to actually connect with Newton's dynamics.⁸³ The parallel is imperfect because chemistry, electricity, magnetism, heat, light were a small number of well-structured complex domains, with relatively weak motivation to achieve unity, while the various areas within cognitive science are many and are much more aware of the need for integration within a respectable field.

So how is the integration supposed to work? The answer lies in the interdisciplinary practices developed over the half-century of cognitive science's existence: practices rather than overarching explicit principles, and interdisciplinary in the sense that any proposal regarding one particular phenomenon under one particular description is offered for discussion to neighboring subfields, and under pressure to achieve compatibility and if possible full articulation with the other descriptions and the other phenomena implicated.⁸⁴ In a sense made familiar by Kuhn and his followers, this is as much a matter of tradition and tacit understanding of what counts as a solution, as an articulation, etc. as one of explicit methodology. But to anyone familiar with the best and most characteristic cognitive science, there is a clear sense of a particular and fruitful search for consilience at work.

However, as we know from democracy, what works with small communities in the early phases of a process may not extend to larger populations over the long term. Two related trends have become apparent in cognitive science in the last decade. One is the exponential growth not only of the field as a whole but of subfields and even subcultures within it, some with strong connections very far from the original core. Thus we have witnessed the appearance of many topics which were not part of the original agenda of cognitive science (consciousness, emotions, self, culture, norms...), and at a higher level of aggregation the rise of the 'affective sciences', 'consciousness studies', 'action theory', and other branches with no fixed name as yet, dealing with rationality, esthetics, ethics, human sociality, culture.... As these new clusters grow, they reach the size and complexity which was that of cognitive science as a whole a mere quarter century back, and necessarily move away from one another to find the breathing space needed for their development. It simply ceases to be feasible, or perhaps even useful, for the individual scientist specializing in one or the other of these new areas to maintain a strong connection with cognitive science as a whole. Perhaps cognitive science is on the verge of becoming an idle structure, like the Roman Empire toward the end, ripe for elimination.

The second trend is the multiplication of proposals which go against LCF, either directly or by implication. In fact, many philosophers of cognitive science these days are occupied with either perfecting, or evaluating the claims of a critical proposal of that sort. Readers familiar with the field may in fact ask why so much of the present chapter is devoted to presenting and discussing LCF, in terms which are not very different from those which would have been appropriate 20 years ago, given that it has been so severely criticized, leaving it, one would think, no other fate than rejection or obsolescence.

It might appear that the only goal worth pursuing at this juncture is to evaluate those attacks and to reach a considered judgment on whether they succeed or fail. Although I agree that it is a goal worth pursuing, I don't believe it is the only one,

and in fact I doubt whether it really is a goal rather than a horizon. My aim in these concluding lines is rather to assess the situation resulting from the historical fact that LCF is battered from all sides yet has not fallen into oblivion nor been replaced.

From the start, LCF came under fire: Dreyfus thought of it as the ultimate expression of the rationalist tradition in philosophy, and thus both highly plausible in the contemporary context and deeply wrong;⁸⁵ Searle has tried to show that it is absurd and based on a massive mistake. A little later, Putnam, who put it on the map of analytic philosophy, rejected it, and by now has spent more time and ink criticizing it than he did setting it up; Chomsky, whose contribution is seen by many as having been decisive for LCF, is critical of the way it is construed today.⁸⁶ Just as dangerous, if not more, for LCF are the indirect threats posed by research programs predicated on assumptions which, if true, would render LCF either false or irrelevant; examples of such programs abound, ranging from the more general (dynamicism, neo-Gestalt and phenomenological approaches, constructivism, neuronalism, radical externalism, and other proposals which either reject or profoundly alter the key LCF concept of mental representation) to the more local (having to do, for example, with emotions, motivation, action, perception, consciousness, memory, norms, the body,...).

Yet LCF still plays a role in the debate. It is a fact that no discussion of the foundations and structure of cognitive science can proceed without first setting the stage, and LCF does exactly that. Nor does it serve as a mere historical preliminary, for if it were the case, LCF could be replaced by its current successor, and there is no successor. Simon and Newell proposed a version of LCF, the 'physical symbol system hypothesis'⁸⁷ and claimed for it the status of a 'law of qualitative structure', comparable to Pasteur's germ theory of disease. But if this is right and the LCF is wrong, then the entire field collapses: if illnesses were not typically caused by germs, then the germ hypothesis would have been rejected and there would remain no such thing as the medicine of infectious diseases; instead, we would have other research programs in medicine, based on our current best theory of the causes of illness.

We have something of a puzzle on our hands, one which may remind philosophers of science of the one posed by the status of 'the Legend', Kitcher's name for a view of science which was by and large adopted by the great thinkers who set up philosophy of science as an academic field (at least in the United States), and has been under such systematic attack for the last 40 years or so that it can no longer pass as an acceptable first shot, to be straightened out by suitable amendments.⁸⁸

In the case of cognitive science and the structuring or foundational role of LCF, there seem to be three main ways of dealing with the puzzle.

The first is to accept Simon and Newell's idea of the physical symbols system hypothesis as a structural hypothesis, but to grant that it has since co-opted many of its opponents' ideas and is now considerably more complex, and looser, than it was. The worry here is: How much structuring can be achieved by an ever looser LCF, one which accepts just about any amendment, caveat and heretical thesis which one can think of? Some defenders of LCF will insist on the other hand that LCF is far less diluted by the amendments and far less threatened by the attacks than an excited lot would like to think.

The second option is to deny that LCF plays any role today, other than loosely delimiting the subject-matter of cognitive science, and that an alternative foundation is gestating. For a long time, I defended the view that LCF was something like a rocket used to launch a satellite (cognitive science), only to be discarded once a certain altitude has been reached. But if this were right, we could talk about cognitive science without referring to LCF just as one can talk, as in fact one does, about contemporary physics without any mention of Descartes' mechanical philosophy, although the latter was instrumental in putting physics on its present track. Note that it may soon appear to be the case, but the necessary recasting of the field, if it is possible, remains to be done.

The third way out is to give up entirely on the need for a structural hypothesis. The usual argument is to compare cognitive science with biology. Biology thrives without the help of a structural hypothesis, and it thrives despite being highly diverse and without any prospect of a grand unification; biology just is the systematic study of a million different phenomena, loosely connected by transverse themes such as the cell, the gene, evolution and the basic features of most living systems (metabolism, growth, homeostasis, etc.). This line however raises two objections. One is that this view of biology as non-unified and devoid of structural principles, however faithful it may be to the present state of the field, is not necessarily a desirable or permanent feature. The other is that while there is little reason to fear that biology will blow up in many fragments and disappear as a conceptual and institutional entity, this is precisely what may be in store for cognitive science, an outcome which motivates in the first place the inquiry on LCF and its possible replacement. Nothing rules out the possibility that the 21st century will feel no more concerned about the project of *a* science of cognition than the 20th was about the conditions of possibility of *a* science of man. We have been living, some happily, others less so, with an institutional label almost entirely devoid of theoretical content, the sciences of man. One sees with increasing frequency references to the cognitive sciences, or the sciences of cognition, the plural indicating an indifference to, or distancing from any pretension to unity in whatever sense. It may turn out that LCF will have been no more than a philosophers' rational reconstruction of an early phase (or perhaps even just a phase *tout court*) in the history of psychology, linguistics and related disciplines.

This question, and the issues leading into it which were barely scratched here, are a large part of the present agenda of philosophy of cognitive science. By dwelling at some length on the formulation of the problem, I hope to have given a sense of what my colleagues are working on now and which will presumably occupy them for years to come. Their conclusions, based in part on, and determining to some extent what path cognitive science will end up following, will be of paramount importance to the sciences of man.

Endnotes

- 1 Including abundant web resources: Stanford Encyclopaedia of Philosophy – <http://plato.stanford.edu/>; Dictionary of Philosophy of Mind – <http://philosophy.uwaterloo.ca/MindDict/>; Internet Encyclopedia Encyclopaedia of Philosophy – <http://www.iep.utm.edu/>; Field Guide in the Philosophy of Mind (Nani, M. & Marraffa, M., eds.) – <http://host.uniroma3.it/progetti/kant/field/>; Chalmers, D., Contemporary Philosophy of Mind: An Annotated Bibliography – <http://consc.net/biblio.html>
- 2 In a highly enlightening and influential essay entitled “The logical geography of computational approaches: a view from the east pole” (1986), Dennett contrasted two radically different conceptions of cognitive science, one cultivated at and near MIT, on the east coast of the United States, the other in California.
- 3 See Andler, 2006a.
- 4 See Piatelli-Palmarini, 1979.
- 5 Even if one were to discount the work done in Continental Europe (which is not acceptable even as an idealization), the UK was very strong from the start. This would make cognitive science and the associated subfield of philosophy a US–UK import on the Continent. Although mistaken, that view is less so than the ‘all US’ theory.
- 6 See Parot et Richelle, 1992; Berthoz, 1999.
- 7 Changeux, 1983, 2003; Berthoz, 1997; Jeannerod, 1997; Imbert, 2006.
- 8 Thom, 1988; Petitot *et al.*, 1999, Some of these ideas started converging in the 1990s with the connectionist and dynamicist views developed in the US and elsewhere (see below).
- 9 For example, the opening sentence of F. Jackson’s and G. Rey’s article “Mind, philosophy of” in Craig (1998) reads as follows: “‘Philosophy of mind’, and ‘philosophy of psychology’ are two terms for the same general area of philosophical inquiry: the nature of mental phenomena and their connection with behaviour and, in more recent discussions, the brain”. M. Davies’ “Cognitive science” in Jackson & Smith (2005) is one chapter of a part entitled “Philosophy of mind and action” (next to “Consciousness”, “Intentionality” and “Action”), and consists almost entirely in a discussion of central topics in philosophy of mind. Note that this distribution of titles makes philosophy of cognitive science a part of philosophy of mind, while others might draw the inclusion sign in the opposite direction, and I prefer to think of the two as distinct areas having a broad intersection. F. Jackson, again, had previously co-authored a book with D. Braddon-Mitchell whose title seems to reflect yet another way of cutting the pie: *Philosophy of Mind and Cognition*.
- 10 Note however that this chapter is not particularly aimed at them, but rather at practitioners of other branches of philosophy of science.
- 11 Dreyfus, 1972; Haugeland, 1998 (original paper 1981), Taylor, 2005 (original paper 1971).
- 12 Andler, 2006b, Smith and Thomasson, 2005.
- 13 An exception was mentioned above, see note 3.
- 14 Alas, I discover only now, as I finalize this chapter, that Paul Livingston’s 2002 dissertation, and his subsequent 2004 book, are devoted to precisely this enterprise, and it seems that his motivation was exactly the worry I express here.
- 15 *E.g.* Rosenthal (1991), Morton (1996), Chalmers (2002), Heil (2004).
- 16 There are exceptions, of course. For example, there are monographs on Kant as a precursor of some key ideas in contemporary cognitive psychology and philosophy of mind: Kitcher, 1990; Brooke, 1994. The limits of such works is that they tend to rehabilitate only one particular line of ancestry (the Kantian one being admittedly of unusual importance).
- 17 Warren S. McCulloch, “What Is a Number, that a Man May Know It, and a Man, that He May Know a Number?”, in McCulloch (1965) (original paper 1961).
- 18 See however Ludwig (2003) for a particularly clear and thorough exposition; Warner & Szubka (1994) is a good collection of papers.
- 19 For a less sanguine view, see Johnson & Erneling, 1997.

- 20 There is in fact an entire tradition of attempts to infer such an impossibility from Gödel's incompleteness theorems, originating in Lucas, 1961.
- 21 I am not referring to the many varieties of functionalism outside psychology (linguistics, anthropology, etc.), although there presumably is an even more abstract core common to all forms of functionalism.
- 22 As Hatfield, 1995 makes clear, some major figures in 18th century psychology had already grasped the possibility of combining a naturalistic approach with an agnostic attitude toward ontological monism (or even an acceptance of dualism).
- 23 I do not discuss at this point the difficulties of functionalism or of any one of its varieties. However it is hard not to notice one glaring threat on psychofunctionalism. If the meaning of B is determined by the set (P) of platitudes of everyday psychology, but the scientific inquiry leads one to the conclusion that B does not obey any approximation of (P), what is one to do? The 'Eddington' move (accepting both the commonsense representation of the table as hard, etc., and the physical representation as a cloud of particles) is not available, as we have no grip on a given belief comparable to the one we have on the table in front of us. We seem forced to eliminate B from our ontology. But if this happens for all, or even for many mental state terms, hasn't the topic vanished altogether? One response is to rule out as highly improbable, or at any rate, unmotivated, this worst-case scenario, and to stress the plausibility of scientific psychology rectifying commonsense platitudes without massively contradicting them. This is the bare beginning of a discussion which has been raging for three decades.
- 24 Putnam, 1975 (the first original paper is dated 1960).
- 25 Putnam, 1988.
- 26 So much so, in fact, that in many contexts (outside 'advanced' philosophy of mind or philosophy of psychology), functionalism just means CTM. (Of course, functionalism means many other things in the context of other disciplines).
- 27 This was not clear, I think, to the ancestors of CTM, viz. the cyberneticians who started working in the 1940s on the question of how a brain could think: see Andler, 1992. For Turing, on the other hand, it was clear, only too clear: he failed to see that representation posed no less a challenge than computation: see Andler, 1998.
- 28 There is, to be sure, a lot more to say regarding the status of X as a token of the X-type: the device needs to 'know' that despite inevitable differences between the X tokens it runs into, they all count as interchangeable tokens of the same type. Thus in a philosophically interesting sense, any X token *represents* type X. Proceeding along this line might yield a resolution of an old controversy between Jerry Fodor and Hilary Putnam, the former holding that 'no representation, no computation', on the grounds that a computation necessarily implies a computational *domain*, and the latter arguing that the theoretical foundation of computation was historically developed, and can be entirely presented, without any reference whatsoever to a notion of representation.
- 29 This is the expression which Putnam 1994 applies to functionalism (or CTM) even *after* it has been beefed up. I will return shortly to the incompleteness of CTM.
- 30 They also argue against putting too much weight on past authors' alleged prescience of the CTM (Ockham, Hobbes, Leibniz, Kant, Boole, Babbage et al.).
- 31 For an appraisal of the significance of the universal Turing machine, see Herken, 1988.
- 32 See the papers by Fodor, Fodor and Pylyshyn, and Smolensky in Horgan and Tienson, 1996 or Macdonald and Macdonald, 1995.
- 33 Dretske, 1981; Millikan, 1984; Papineau, 1987.
- 34 This is non-technical shorthand for a notion of logic which depends on the logical system one believes to be the correct one for LoT.
- 35 This may seem unfair and unwarranted: some might dispute the failure of early AI, and, even if one grants it, to incriminate this particular assumption among the many initial theoretical and technological conditions which presided over the genesis of AI requires some justification. Hofstadter, 1985: Chap. 26, pp. 631—665 is a detailed attempt to do just that.
- 36 I owe the references to Hamilton and to Bain to the above-mentioned chapter by Martin Davies (Davies 2005).

- 37 Bain, 1893, p. 48, quoted in Davies, 2005.
- 38 Chomsky, 1965.
- 39 Dennett, 1969.
- 40 Hofstadter, 1985: Chap. 26.
- 41 Rumelhart & McClelland, 1986.
- 42 Smolensky, 1988 (reprinted in refs given in note 31). For a more recent presentation of his view, see Smolensky & Legendre, 2006, vol. 2, Chap. 23.
- 43 Bermudez, 2003.
- 44 Dreyfus, 1972; Searle, 1992.
- 45 Dennett, 1969.
- 46 On priming: Marcel, 1983; on blindsight: Weiskrantz, 1986; on amnesia: Warrington & Weiskrantz, 1968. For more details see e.g. Frith and Rees in Velmans and Schneider, 2007 (chap. 1).
- 47 The *locus classicus* is Nagel's famous 1974 paper; Levine, 1983 coined the felicitous expression 'explanatory gap' to refer to what separates the target of any conceivable scientific account of consciousness from actual conscious experience.
- 48 Block, 1995.
- 49 Chalmers, 1996, and in Velmans and Schneider, 2007 (chaps.17 and 28).
- 50 See Velmans and Schneider, 2007, in particular the list of institutions on pp. 727—728.
- 51 See e.g. Levine in Velmans and Schneider, 2007 (chap. 29).
- 52 Unfortunately space does not allow for even a preliminary discussion of the distortions affecting many historical accounts of scientific psychology. The issue raised in the text depends of course on getting one's history reasonably straight. The reader is referred to Gary Hatfield's pioneering revisionary studies (Hatfield, 1995, 1997, 2002).
- 53 I am leaving aside the difficult matter of deciding whether there is a useful notion of computation which goes beyond the classical theory of recursive (Turing-computable) functions *and* is relevant for cognitive scientific models.
- 54 The metaphor respects relatively obvious features of the original: a pick-up truck cannot carry mount Everest or the Gange, it cannot cross oceans, etc. The mind balks at calculating the 10000th decimal of pi, stops short of parsing a sentence a million words long, and cannot deploy a winning strategy for chess, despite the fact that it exists. Technically, this sort of limitation is abundantly studied under the rubric of complexity or feasibility. I personally remain unconvinced that it is worth the effort, but mine is definitely a minority view. We will return to this theme in §3.b below.
- 55 Piatelli-Palmarini, 1979.
- 56 Fodor, 1983.
- 57 D. Sperber *in* Dupoux, 2001; Hirschfeld & Gelman, 1994; Barkow et al. 1992; Carruthers et al. 2005.
- 58 Sperber, 2005 proposes an answer.
- 59 See e.g. the first section of Stainton, 2006.
- 60 See Carruthers et al. 2005.
- 61 See e.g. B. Scholz & G. Pullum, *in* Stainton, 2006, chap.4; Cowie, 1999.
- 62 Elman et al., 1996; Quartz & Sejnowski, 1997.
- 63 Compare with the disputes which the notion of gene has occasioned in the last quarter century: they have not resulted in throwing the entire domain of the life sciences, or even of genetics, in disrepute.
- 64 Cosmides, 1989.
- 65 Johnson-Laird, 1983.
- 66 Rips, 1986, Bonatti, 1994; Andler, 1995.
- 67 Johnson-Laird, 1988.
- 68 Simon, 1957.
- 69 Cozic, 2005 for a careful analysis of the notion of bounded rationality.
- 70 Gigerenzer & Selten, 2002.
- 71 Stich, 1996 and later publications. See also Nisbett, 2003.
- 72 Snyder & Lopez, 2002.

- 73 Smolensky, 1988; Smolensky and Legendre, 2005. Smolensky was a member of the PDP Group.
- 74 For another, equally important and somewhat different version of connectionism, see Amit, 1989. An illuminating philosophical account of connectionism is provided in Clark, 1989.
- 75 "A logical calculus of ideas immanent in nervous activity", in McCulloch, 1965; see Anderson and Rosenfeld, 1988; Andler, 1992.
- 76 McDonald & McDonald, 1995 or Horgan & Tienson, 1996.
- 77 See refs. in note 7. Andler, 1990.
- 78 See refs. in note 73.
- 79 Simon, 1979.
- 80 Andler, 2005; Ravenscroft, 1998, or for an all-out attack on the explanatory pretensions of cognitive neuroscience, Bennett and Hacker, 2003.
- 81 Note that naturalism need not subscribe to eliminativism.
- 82 Defended in particular in Bickle, 2003.
- 83 Harman, 1982.
- 84 This is shown inter alia by the fact that one does not often find in the cognitive science literature immunizing clauses such as '*X as understood in the field I work in*', which prevent anyone looking at X from another angle to object. This is in stark contrast with a large majority of research traditions in the human and social sciences. Note however that generative linguistics and its subfields make extensive uses of such clauses. This is not the place to ask why.
- 85 Interestingly, Dreyfus's critique of the radical version of rationalism offered by AI and the then nascent cognitive psychology echoes Neurath's attack on 'hyperrationalism' (Neurath, 1983; original papers from 1922).
- 86 Dreyfus, 1972; Searle, 1992; Putnam, 1988; Chomsky, 2000, 2002.
- 87 Newell and Simon, 1976.
- 88 Kitcher, 1995.

Bibliography

- Amit Daniel (1989), *Modeling Brain Function — The World of Attractor Neural Networks*, Cambridge, Cambridge University Press, 1989.
- Anderson James A. and Rosenfeld Edward (1988) (eds.), *Neurocomputing. Foundations of Research*, Cambridge, MA, MIT Press.
- Andler Daniel (1990), "Connexionnisme et cognition: à la recherche des bonnes questions", *Revue de Synthèse*, série générale CXI, 1–2, Janv.–Juin 1990, pp. 95–127.
- Andler Daniel (1992), "From paleo to neo-connectionism", in G. van der Vijver (ed.), *New Perspectives on Cybernetics*, Kluwer, Dordrecht, pp. 125–146.
- Andler Daniel (1995), "Logique, raisonnement et psychologie", in J. Dubucs and P. Lepage (eds.), *Méthodes logiques pour les sciences cognitives*, Paris, Hermes, pp. 25–75.
- Andler Daniel (1998), "Turing: Pensée du calcul, calcul de la pensée", in F. Nef and D. Vernant, *Les années 1930: réaffirmation du formalisme*, Paris, Vrin.
- Andler Daniel (2004) (ed.), *Introduction aux sciences cognitives*, 2nd ed., Paris, Gallimard.
- Andler Daniel (2005), "Les neurosciences cognitives: une nouvelle 'nouvelle science de l'esprit'?", *PSN – Psychiatrie, sciences humaines, neurosciences*, 3(12), pp. 74–87.
- Andler Daniel (2006a), "Cognitive Science", in L. Kritzman (ed.), *The Columbia History of Twentieth Century French Thought*, New York, Columbia University Press, pp. 175–181.
- Andler Daniel (2006b), "Phenomenology in cognitive science and artificial intelligence", in H. Dreyfus and M. Wrathall (eds.), *A Companion to Phenomenology and Existentialism*, Oxford, Blackwell, pp. 377–393.
- Bain A. (1893), "The respective spheres and mutual helps of introspection and psychophysical experiment in psychology", *Mind*, 2, pp. 42–53.

- Barkow Jerome, Cosmides Leda and Tooby John (1992) (eds.), *The Adapted Mind. Evolutionary Psychology and the Generation of Culture*, New York, Oxford University Press.
- Bechtel W. and G. Graham (1999) (eds.), *A Companion to Cognitive Science*, Oxford, Blackwell.
- Bennett M.R. and P.M.S. Hacker (2003), *Philosophical Foundations of Neuroscience*, Oxford, Blackwell.
- Bermúdez J. L. (2003), *Thinking Without Words*, Oxford, Oxford University Press.
- Bermúdez J. L. Marcel A., Eilan N. (1995) (eds.), *The Body and the Self*, Cambridge, Massachusetts, MIT Press.
- Berthoz Alain (1997), *Le sens du mouvement*, Paris, Odile Jacob. *The Brain's Sense of Movement*, Cambridge, MA, Harvard University Press, 2002.
- Berthoz Alain (1999) (ed.), *Leçons sur le corps, le cerveau et l'esprit*, Paris, Odile Jacob.
- Bickle J. (2003), *Philosophy and Neuroscience. A Ruthlessly Reductive Account*, Dordrecht, Kluwer.
- Block Ned (1995), "On a Confusion about a Function of Consciousness", *The Behavioral and Brain Sciences*, 18, 2, p. 1995, pp. 227–247. Reprinted in Flanagan and Güzeldere, 1997, Block, 2007.
- Block Ned (2007), *Consciousness, Function, and Representation. Collected papers*, vol. 1, Cambridge, MA, MIT Press.
- Bonatti Luca (1994), "Propositional Reasoning by Model?", *Psychological Review*, 101(4), pp. 725–733.
- Braddon-Mitchell D. and Jackson F. (1996), *Philosophy of Mind and Cognition*, Oxford, Blackwell.
- Brook Andrew (1994), *Kant and the Mind*, Cambridge, Cambridge University Press.
- Carruthers Peter, Laurence Stephen and Stich Stephen (2005) (eds.), *The Innate Mind, 1: Structure and Content*, New York, Oxford University Press.
- Chalmers David (1996), *The Conscious Mind*, New York, Oxford University Press.
- Chalmers David (2002) (ed.), *Philosophy of Mind: Classical and Contemporary Readings*, Oxford, Oxford University Press
- Changeux Jean-Pierre (2003), *L'homme de vérité*, Paris, Odile Jacob,
- Changeux Jean-Pierre (1983), *L'homme neuronal*, Paris, Fayard. *Neuronal Man: The Biology of Mind*, trans. Laurence Garey, Princeton, Princeton University Press, 1997.
- Chomsky Noam (1965), *Aspects of the Theory of Syntax*, Cambridge, MA, MIT Press.
- Chomsky Noam (2000), *New Horizons in the Study of Mind and Language*, Cambridge, Cambridge University Press.
- Chomsky Noam (2002), *On Nature and Language*, Cambridge, Cambridge University Press.
- Clark Andy (1989) *Microcognition. Philosophy, Cognitive Science, and Parallel Distributed Processing*, Cambridge, Massachusetts, MIT Press.
- Cosmides L. (1989), "The Logic of Social Exchange: Has natural selection shaped how humans reason? Studies with the Wason selection task", *Cognition*, 31, pp. 187–276.
- Cowie F. (1999), *What's Within? Nativism Reconsidered*, Oxford, New York, Oxford University Press.
- Cozic Mikaël (2005), "Fondements cognitifs du choix en rationalité limitée", PhD dissertation, Université de Paris-Sorbonne (Paris IV).
- Craig E. (1998) (ed.), *Routledge Encyclopedia of Philosophy*, London, Routledge.
- Davidson R.J., Scherer K.R. and Goldsmith H.H. (2003) (eds.), *Handbook of Affective Sciences*, Oxford, Oxford University Press.
- Davies Martin (2005), "Cognitive science", in F. Jackson and M. Smith (eds.), 2005, pp.358–394.
- Dennett Daniel C. (1969), *Content and Consciousness*, London, Routledge and Kegan Paul.
- Dennett Daniel C. (1978), *Brainstorms*, Montgomery, VT, Bradford Books.
- Dennett Daniel C. (1986), "The Logical Geography of Computational Approaches", in M. Brand and R. M. Harnish, (eds.), *The Representation of Knowledge and Belief*, Tucson, University of Arizona Press, pp. 59–79.
- Dretske Fred (1981), *Knowledge and the Flow of Information*, Cambridge, Massachusetts, MIT Press.

- Dreyfus Hubert L. (1972), *What Computers Can't Do*, New York, Harper & Row; augm. edition: *What Computers Still Can't Do*, Cambridge, MA, MIT Press, 1993.
- Dreyfus Hubert L. and Harrison Hall (1982) (eds.), *Husserl, Intentionality, and Cognitive Science*, Cambridge, Massachusetts, MIT Press.
- Dupoux Emmanuel (2001) (eds.), *Language, Brain and Cognitive Development: Essays in Honor of Jacques Mehler*, Cambridge, MA, MIT Press.
- Elman J., Bates L.E., Johnson M.H., Karmiloff-Smith A., Parisi D. and Plunkett D. (1996), *Rethinking Innateness. A Connectionist Perspective on Development*, Cambridge, Massachusetts, MIT Press.
- Enfield N.J., Levinson Stephen and Stephen C. (2006) (eds.), *Roots of Human Sociality. Culture, Cognition and Interaction*, Oxford, Berg.
- Flanagan O. and G. Güzeldere (eds.) (1997), *The Nature of Consciousness*. Cambridge, Massachusetts, MIT Press.
- Fodor Jerry (2000), *The Mind Doesn't Work that Way. The Scope and Limits of Computational Psychology*, Cambridge, Massachusetts, MIT Press.
- Fodor Jerry (1983), *The Modularity of Mind*, Cambridge, Massachusetts, MIT Press.
- Fodor Jerry (1975), *The Language of Thought*, New York, Thos. Crowell; repr. Cambridge, MA, Harvard University Press.
- Gardner H. (1985), *The Mind's New Science*, New York, Basic Books.
- Gazzaniga M. (2000), *The New Cognitive Neurosciences*, 2nd ed., Cambridge, Massachusetts, MIT Press.
- Gigerenzer Gerd and Reinhard Selten (2002) (eds.), *Bounded Rationality: The Adaptive Toolbox*, Cambridge, Massachusetts, MIT Press.
- Guttenplan Samuel (1994) (ed.), *A Companion to the Philosophy of Mind*, Oxford & Cambridge, Massachusetts, Blackwell.
- Harman P.M. (1982), *Energy, Force, and Matter*, Cambridge, Cambridge University Press.
- Hatfield Gary (1995), "Remaking the science of the mind. Psychology as natural science", in C. Fox, R. Porter and R. Wokler (eds.), *Inventing Human Science. Eighteenth-Century Domains*, Berkeley & Los Angeles, University of California Press.
- Hatfield Gary (1997), "Wundt and psychology as science: Disciplinary transformations", *Perspectives on Science*, 5(3), pp. 349–382.
- Hatfield Gary (2002), "Psychology, philosophy, and cognitive science: Reflections on the history and philosophy of experimental psychology", *Mind and Language*, 17(3), pp. 207–232.
- Haugeland John (1998), *Having Thought*, Cambridge, Massachusetts, Harvard University Press.
- Heil John (2004) (ed.), *Philosophy of Mind: A Guide and Anthology*, New York, Oxford, Oxford University Press.
- Herken R. (1988) (ed.), *The Universal Turing Machine. A Half-Century Survey*, Oxford, Oxford University Press.
- Hirschfeld Lawrence A. and Gelman Susan A., (1994) (eds.), *Mapping the Mind. Domain Specificity in Cognition and Culture*, Cambridge, Cambridge University Press.
- Hofstadter Douglas (1985), *Metamagical Themas*, New York, Basic Books.
- Horgan T. and Tienson J. (1996) (eds.), *Connectionism and the philosophy of psychology*, Cambridge, Massachusetts, MIT Press.
- Imbert Michel (2006), *Traité du cerveau*, Paris, Odile Jacob.
- Jackson F. and M. Smith, (eds.) (2005), *The Oxford Handbook of Contemporary Philosophy*, Oxford, Oxford University Press.
- Jacob Pierre (1997), *What Minds Can Do*, Cambridge, Cambridge University Press.
- Jeannerod Marc (1997), *The Cognitive Neuroscience of Action*, Cambridge, Cambridge University Press.
- Johnson David and Erneling Christina (eds.) (1997), *The Future of the Cognitive Revolution*, New York, Oxford, Oxford University Press.
- Johnson-Laird Philip N. (1983), *Mental Models*, Cambridge, Cambridge University Press.
- Johnson-Laird Philip N. (1988), *The Computer and the Mind. An Introduction to Cognitive Science*, Cambridge, Massachusetts, Harvard University Press.

- Kahneman D., Slovic P. and Tversky A. (1982), *Judgment Under Uncertainty: Heuristics and Biases*, Cambridge, Cambridge University Press.
- Kitcher Patricia (1990), *Kant's Transcendental Psychology*, New York, Oxford, Oxford University Press.
- Kitcher Philip (1995), *The Advancement of Science: Science Without Legend, Objectivity Without Illusions*, New York, Oxford, Oxford University Press.
- Levine J. (1983), "Materialism and Qualia: The Explanatory Gap", *Pacific Philosophical Quarterly*, 64, pp. 354–361.
- Livingston Paul (2004), *Philosophical History and the Problem of Consciousness*, Cambridge, Cambridge University Press.
- Lucas J.R. (1961), "Minds, Machines and Gödel", *Philosophy*, 36, pp. 112–127; repr. in Alan R. Anderson (ed.), *Minds and Machines*, Prentice Hall, 1964.
- Ludwig K. (2003), "The Mind-Body Problem: An Overview," in Warfield & Stich, 2003, pp. 1–46.
- Macdonald Cynthia and Macdonald Graham (1995) (eds.), *Connectionism*, London, Blackwell.
- McCulloch Warren S. (1965), *Embodiments of Mind*, Cambridge, Massachusetts, MIT Press.
- Marcel A. J. (1983), "Conscious and Unconscious Perception: Experiments on Visual Masking and Word Recognition", *Cognitive Psychology*, 15, pp. 197–237.
- Millikan Ruth (1984), *Language, Thought and Other Biological Categories*, Cambridge, Massachusetts, MIT Press.
- Morton P. (1996) (ed.), *A Historical Introduction to the Philosophy of Mind: Readings with Commentary*, Calgary, Alberta, Broadview Press.
- Nadel L. (2003) (gen. ed.), *Encyclopaedia of Cognitive Science*, London, New York & Tokyo, Nature Publishing Group.
- Nagel T. (1974), "What Is It Like To Be a Bat?", *Philosophical Review*, 83(4) pp. 435–450.
- Neurath Otto (1983), *Philosophical Papers, 1913–1946*, trans. Robert S. Cohen and Marie Neurath (eds.), Dordrecht, D. Reidel.
- Newell Allen and Simon Herbert (1976), "Computer Science as Empirical Inquiry: Symbols and Search", *Communications of the ACM*, 19(3), pp. 113–126.
- Nisbett Richard E. (2003), *The Geography of Thought*, New York, The Free Press.
- Papineau David (1987), *Reality and Representation*, Oxford, Basil Blackwell.
- Parot Françoise and Richelle Marc (1992) (eds.), *Psychologues de langue française*, Paris, PUF.
- Petisstot J., Varela F.J., Pachoud B. and Roy J.M. (1999) (eds.), *Naturalizing Phenomenology: Issues in Contemporary Phenomenology and Cognitive Science*, Stanford, Stanford University Press.
- Piattelli-Palmarini Massimo (1979) (ed.), *Théories du Langage, Théories de l'Apprentissage, le Débat Chomsky/Piaget*, Paris, Seuil; *Language and Learning: The debate between Jean Piaget and Noam Chomsky*, London, Taylor & Francis, 1983.
- Port R.F. and van Gelder T. (1995) (eds.), *Mind as Motion: Explorations in the Dynamics of Cognition*, Cambridge, Massachusetts, MIT Press.
- Putnam Hilary (1960), "Minds and Machines", reprinted in Putnam, 1975.
- Putnam Hilary (1975), *Mind, Language, and Reality*, Cambridge, Cambridge University Press.
- Putnam Hilary (1988), *Representation and Reality*, Cambridge, Massachusetts, MIT Press.
- Putnam Hilary (1994), Entry "Putnam, Hilary" in Guttenplan (1994).
- Quartz Steven R. and Sejnowski Terrence J. (1997), "The Neural Basis of Cognitive Development: A Constructivist Manifesto", *Behavioral and Brain Sciences*, 20(4), pp. 537–596.
- Ravenscroft Ian (1998), "Neuroscience and the mind", *Mind and Language*, 13, p. 132–137.
- Rips L.J. (1986), "Mental Muddles", in M. Brand & M. Harnish, eds, *The Representation of Knowledge and Belief*, Tucson, University of Arizona Press, pp. 258–286.
- Rosenthal David M. (1991) (ed.), *The Nature of Mind*, New York, Oxford University Press.
- Rowlands Mark (2003), *Externalism*, Montreal & Kingston, McGill-Queens University Press.
- Rumelhart D., McClelland J. and the PDP Research Group (1986), *Parallel Distributed Processing. The Microstructure of Cognition*, Cambridge, Massachusetts, MIT Press.
- Searle John R. (1992), *The Rediscovery of Mind*, Cambridge, Massachusetts, MIT Press.

- Simon Herbert (1957), "A Behavioral Model of Rational Choice", in *Models of Man, Social and Rational: Mathematical Essays on Rational Human Behavior in a Social Setting*, New York, Wiley.
- Simon T. W. (1979), "Philosophical Objections to Programs as Theories", in M. Ringle (ed.), *Philosophical Perspectives in AI*, Atlantic City, NJ, Humanities Press.
- Smith David W. and Thomasson Amie L. (2005) (eds.), *Phenomenology and Philosophy of Mind*, New York, Oxford University Press.
- Smolensky Paul (1988), "On the proper treatment of connectionism", *Behavioral and Brain Sciences*, 11, pp. 1–74.
- Smolensky Paul and Legendre Géraldine (2005), *The Harmonic Mind: From Neural Computation to Optimality-Theoretic Grammar*, vol. 1: *Cognitive Architecture*; vol. 2: *Linguistic and Philosophical Implications*, Cambridge, Massachusetts, MIT Press.
- Snyder Charles R. and Lopez Shane J. (2002) (eds.), *Handbook of Positive Psychology*, New York, Oxford University Press.
- Sperber Dan (1996), *La Contagion des idées*, Paris, Odile Jacob; *Explaining Culture: A Naturalistic Approach*, Oxford, Blackwell, 1996.
- Sperber Dan (1994), "The modularity of thought and the epidemiology of representations", in Hirschfeld and Gelman, 1994.
- Sperber Dan (2005), "Modularity and Relevance: How can a Massively Modular Mind be flexible and Context-Sensitive?", in Carruthers, Laurence and Stich, 2005.
- Stainton Robert J. (2006) (ed.), *Contemporary Debates in Cognitive Science*, Oxford, Blackwell.
- Stich Stephen (1996), *Deconstructing the Mind*, New York, Oxford University Press.
- Taylor Charles (2005), *Philosophical Papers*, Cambridge, Cambridge University Press.
- Thom René (1988), *Esquisse d'une sémiophysique*, Paris, Interéditions.
- Tomasello Michael (2001), *The Cultural Origins of Human Cognition*, Cambridge, Massachusetts, Harvard University Press.
- Turing Alan M. (1950), "Computing machinery and intelligence", *Mind*, LIX(236), pp. 433–460.
- Velmans Max, Schneider Susan (2007) (eds.), *The Blackwell Companion to Consciousness*, Oxford, Blackwell.
- Warfield T. S. and Stich S. (2003) (eds.), *The Blackwell Guide to the Philosophy of Mind*, Oxford, Blackwell.
- Warner Richard and Szubka Tadeusz (1994) (eds.), *The Mind-Body Problem*, Oxford, Blackwell.
- Warrington E.K. and Weiskrantz L. (1968), "New method of testing long-term retention with special reference to amnesic patients", *Nature*, 217, pp. 972–974.
- Weiskrantz L. (1986), *Blindsight: A Case Study and Implications*, Oxford, Oxford University Press.
- Wilson R. A. and Keil F.C. (1999) (eds.), *The MIT Encyclopedia of the Cognitive Sciences*, Cambridge, Massachusetts, MIT Press.

Part VI
Philosophy of Economics

Duhemian Themes in Expected Utility Theory

Philippe Mongin

1 Introduction

According to the philosophical position of *epistemological holism*, the statements of the empirical sciences do not relate to observations singly, but collectively. This is because these statements belong to logically complex theoretical structures, which are to a large extent indivisible, and also because further theoretical assumptions (an “observational theory”) underlie the observations made to check them empirically. As a consequence of this basic claim, all brands of epistemological holism include an *underdetermination thesis*, to the effect that the scientists’ decisions about hypotheses are underdetermined by the evidence available to them, and in particular, by the results of the tests they perform. Pragmatic reasons must eventually prevail in the choice of attributing the evidence to this or that part of the theoretical whole, and when elaborating on these reasons, philosophers of science will never offer more than partial and context-dependent guidelines.

Duhem’s book, *La théorie physique. Son objet, sa structure* (1906, 2nd ed. 1914), henceforth *TP*, stands out as a landmark for the overall position just sketched. The stunning success of this work is largely due to the way in which Quine expanded on it in the celebrated second chapter of *From a Logical Point of View* (1953, II, 6), henceforth *FLPV*. In the brand of epistemological holism favoured by Quine at the time, the set of statements relating to an observation is nothing short of the entire body of accepted knowledge. The accompanying underdetermination thesis claims that the scientists’ response to empirical evidence may involve reconsidering any elements whatever of that large set, including the mathematical or logical ones. Quine goes on to suggest that the choice between acceptance and rejection is one of mere convenience, and eventually sketches a picture of science as trying to minimize the amount of disturbance brought about by new empirical evidence.

P. Mongin (✉)
GREG-HEC
1 rue de la Libération
78350 Jouy-en-Josas, France
E-mail: mongin@hec.fr

As it turns out, Duhem is much less extreme than Quine in the extent of his epistemological holism. He does not relate physical observations to theoretical statements outside physics, and he typically takes chunks of this discipline to be the collective units of relevance; only exceptionally does he claim that the whole of physics is involved. Consistently, his underdetermination thesis is less dramatic than Quine's, and when he suggests pragmatic remedies, he is both more specific and more convincing than the latter in the wildly speculative conclusion of *FLPV*. Philosophers of science have recognized this difference for some time already, and the post-Quinean catch – the “Duhem-Quine thesis” used to refer to the underdetermination thesis of the two authors at the same time – is not so common today as it once was.¹

To disentangle Duhem and Quine further, we may add that they pursue different strategic aims, and that this disanalogy matters no less than the difference in degree now recognized in the literature. Quine's later work most clearly, but arguably also *FLPV*, gives a *semantic* turn to a position which, for Duhem, was basically *methodological* in character. As is well-known, *FLPV* uses the underdetermination thesis to rebut a dogma of logical positivists – statements can be verified or refuted individually – that underlies their theory of meaning. Although this challenge also matters to methodology, it is not the connection that Quine brings to the fore. Beside the positivist theory of meaning, he wants to debase the further – in his peculiar view, related – dogma that a distinction can be drawn between analytic and synthetic statements, and this is again a semantic target, not a methodological one. Duhem sometimes comes close to discussing meaning issues,² but his main purpose is to explain how physicists can make the best out of empirical findings that never warrant indisputable conclusions. We are committed here to a strictly methodological perspective, and will deal only with Duhem's position, leaving aside both Quine's and the irrelevant compound of the early literature. Correspondingly, when we write “holism” in the sequel, we mean epistemological or methodological, not semantic holism. We are aware of the need of clarifying this distinction, but this is not a task for the present paper.³

Following a deep-rooted French tradition running from Bernard to 20th century philosophers like Bachelard and Canguilhem, Duhem developed a local philosophy of science. He did not mean to extend his conclusions beyond the only science he investigated, which is physics. However, his conclusions can be generalized because they depend on an implicit abstract notion of what a fully grown scientific theory is. The laws of physics, Duhem writes, are both “symbolic” and “approximate”. This can be taken to mean that the laws of physics are formalized, and that at least some of the concepts in the formalized laws can be measured empirically, though always to a degree of approximation. Duhem claim that the laws of physiology have not reached the stage of symbolism and approximation, which leads him to the intriguing conclusion that they evade the problems created by the underdetermination thesis.⁴ Using Duhem's demarcation for what it is worth, we could carry his epistemological holism beyond physics, though not throughout empirical science. The present paper is concerned with applying it to *decision theory*, where some of the generalities counting as putative laws are both “symbolic” and “approximate” in the above sense. Specifically, the paper tries to reinterpret *expected utility theory* in the light of

Duhem's epistemological holism, and it will immediately appear that this particular theory satisfies the two conditions unproblematically. Thus, even on Duhem's terms, the comparison to come is not unwarranted.⁵

Here is an outline of the overall argument. Section 2 discusses two different *Duhem theses* that can be read in *TP*, i.e., the no-crucial-experiment thesis – crucial experiments, even refuting ones, are impossible – and the underdetermination thesis proper – there is no compelling way of selecting which of the substantial hypotheses, auxiliary hypotheses, and even observation statements, is responsible for an empirical refutation. The distinction of these two theses, which is now widely agreed on, casts further light on Duhem's holism and facilitates comparison with the falsificationist construals of underdetermination – especially in Popper (1963) and Lakatos (1970) – which we will also address. Most of the paper is concerned with the problem resulting from the underdetermination thesis, i.e., how can an empirical refutation be apportioned across the list of scientific statements so as to licence the rejection of some and the acceptance of others? However, the no-crucial-experiment thesis will also play some rôle in the analysis.

Philosophers of science have been more impressed by the vivid way in which Duhem raises and exemplifies the underdetermination problem than by the few hints he makes to answer it. But the persistent failure of more precise solutions – which we will document⁶ – has suggested to us that we may fruitfully revisit his suggestions. From *TP* we extracted the following view: although Duhem's underdetermination problem cannot be resolved in a compelling way, it did receive a *satisfactory* solution sometimes in the history of science, and these favourable cases always involved a lengthy process of accumulation and evaluation of conflicting arguments in the face of the evidence, with the arguments typically pointing out in the same direction after a while. Whoever takes this view seriously has no choice but to analyze historical examples painstakingly, as Duhem did in *TP* with his classic account of the theories of light.⁷ For the underlying claim is that a satisfactory solution can only be recognized *from the pattern of the arguments proposed over a sufficiently long period of time*, and this imposes on the philosopher of science the task of considering the successive stages of his object-theory, and not just its final or received state. It is this connection precisely – and not just the familiar point that historical examples are useful in clarifying and supporting philosophy of science claims – which has prompted the case study of this paper.

Accordingly, Sections 3 and 4 reconstruct part of expected utility theory (EUT) with a view of submitting it to the Duhemian grid. These two sections are mostly technical and historical, with the main assessment awaiting the final Section 5. We have skipped many details and actually restricted attention to the simpler of the two branches of EUT – the von Neumann-Morgenstern (VNM) theory of risky choice, which takes probabilities for granted instead of deriving them, as in Savage's more sophisticated contribution. Section 3 starts with *Theory of Games and Economic Behavior* (1944–1947), and Section 4 finishes with the experimental work of the 1980–1990, when the utility theory originating in this book appeared to be superseded by more promising alternatives. Rather than von Neumann and Morgenstern themselves, we emphasize their immediate followers, like Marschak, who gave their work its final axiomatic touch. Unsurprisingly to those acquainted with the

field, the turning point of our narrative is the attack launched in 1952 by Allais against VNM theory and EUT generally, which could then rely on champions of no lesser fame than Marschak, Friedman, Savage and Samuelson. We discuss the major choice experiments that were made in the 1970s and the early 1980s after Allais's suggestions, before we move to some of the burgeoning theoretical developments of the time. This is inevitably very selective, and the interested reader is referred to the excellent surveys available in the field both then and now.⁸ The drift of a long story made short is that VNM theory became swamped with both systematic counterexamples – “effects” in decisiontheoretic parlance – and alternative hypotheses that went some way towards accounting for these counterexamples, a situation which naturally suggests trying a Duhemian account of the whole sequence.

Section 5 first discusses the sense in which VNM theory can be said to be refuted by the choice experiments. This conclusion is far from obvious given the wide range of possibilities for empirical refutation that the narrative brings out. There is the further complication that a good deal of the debate, including Allais's critique, was concerned with the *normative* problem of capturing individual rationality under risk suitably. Despite the obstacles, we will conclude that a Duhemian philosopher could endorse the conclusion in view of the pattern over time of the conflicting arguments. The section also investigates the way in which decision theorists disentangled the possible sources of refutation *within* the theory, i.e., between the axioms, and what rôle exactly was played by surrounding non-VNM hypotheses in the various attempts at providing a solution. Two major results of the inquiry are, for one, that more effort could and should have been done to identify the axiom responsible for the refutation, and for another, that the other hypotheses did not condition the various Duhemian decisions in any important sense. We eventually reject the Lakatosian claim that a genuine refutation requires a theoretical alternative to be available.

As a bibliographical aside, the author may perhaps mention that this is not the first time that he has grappled with Duhemian themes in expected utility theory. Mongin (1988) explored them already, and his paper was actually one the first sustained attempts at discussing holism and underdetermination in the context of – broadly speaking – economics.⁹ This early work has since been superseded, both by ensuing changes in the object-theory and a more refined philosophical perception of it. Here, we catch up with these developments, but do not extend the study beyond its initial time limits. The better hindsight on the same 40-year sequence is sufficient to make the account significantly novel.

2 Duhem's Theses

2.1 A Statement of Duhem's Underdetermination Thesis

The following passage from *TP* (II, Chap. VI, § II, p. 280–281; English trans., p.185) is an elaborate statement of the underdetermination thesis:

A physicist decides to demonstrate the inaccuracy of a proposition; in order to deduce from this proposition the prediction of a phenomenon and institute the experiment which is to show whether this phenomenon is or is not produced, in order to interpret the results of this experiment and establish that the predicted phenomenon is not produced, he does not confine himself to making use of the proposition in question; he makes use also of a whole group of theories accepted by him as beyond dispute. The prediction of the phenomenon, whose non production is to cut off debate, does not derive from the proposition challenged if taken by itself but from the proposition at issue joined to that whole group of theories; if the predicted phenomenon is not produced, not only is the proposition questioned at fault, but so is the whole theoretical scaffolding used by the physicist. The only thing the experiment teaches us is that among the propositions used to predict the phenomenon and to establish whether it would be produced, there is at least one error; but where this error lies is just what it does not tell us. The physicist may declare that this error is contained in exactly the proposition he wishes to refute, but is he sure it is not in another proposition? If he is, he accepts implicitly the accuracy of all the other propositions he has used, and the validity of his conclusion is as great as the validity of his confidence.

A striking feature of this passage is that it describes the experimental physicist as wishing to achieve refutations, not verifications, nor even confirmations – if, as many do, one takes this third notion to be a weakening of the second. Perusal of *TP* suggests that Duhem would accept that much of Popper's philosophy: *empirical evidence informs the scientist via refutations*. Consistently with this stand, Duhem's underdetermination thesis can only mean that an informative piece of evidence undermines a certain conjunction of statements among which it does not help to identify the false ones. This quasi-Popperian reading of *TP* seems to have been accepted nearly universally – albeit tacitly, which comes as a surprise.¹⁰

The selected passage conveys a variety of reasons to conclude that scientific theories are underdetermined by empirical evidence, in the refutationist sense just stated. It argues that physicists must hypothesize a large number of premisses in order to deduce a test statement, i.e., one which can be compared logically with those of an observational record. If this statement clashes with the observational record, one of the premisses must be false, and this is all one can conclude from propositional logic. Within the premisses, there is a natural distinction to be drawn, even if it is not recorded as such by Duhem, between *primary* and *auxiliary* hypotheses. The former belong to the theory under investigation, and typically reappear from one experiment – more generally, one empirical test – to another. Hypotheses of the latter group must be added if the test statement is to be deduced from the former, and they are often required only for the test at hand. But this is not all there is to Duhem's underdetermination thesis. Although this may seem subdud in the passage, he is not willing to consider the observational record as any more secure than the premisses from which the test statement conflicting with this record is derived. So we should conclude that for Duhem, the set of scientific statements creating the indeterminacy is really *threefold*: (i) primary hypotheses, (ii) auxiliary hypotheses, and (iii) the observational record. Each item can be recognized in the second sentence: "in order to deduce from this proposition the prediction of a phenomenon and institute the experiment which is to show whether this phenomenon is or is not produced".

Without doing full justice to the complex historical examples in *TP*, this scheme fits with them acceptably well. When discussing Foucault's celebrated experiment on the speed of light, Duhem only mentions statements contained in group (i), but his account of another experiment in optics, Wiener's – which was intended to refute an earlier significant hypothesis by Neumann – underscores the relevance of groups (ii) and (iii).¹¹ By analogy with this example, Foucault's experiment can be analyzed in terms of the three categories. Here, (i) stands for the corpuscular theory of light, (ii) for the various assumptions Foucault had to make in order to measure the velocity of light in water and in the air, and (iii) for the observational record of the respective velocities, once the experiment was performed. Recall that the corpuscular theory entailed the putative regularity that light was *faster* in water than in the air. Foucault's experimental setting, whereby the speed of light could be measured in both environments, turned this statement into a test prediction, which apparently failed.

The threefold distinction can be refined if necessary. For instance, if one is concerned with explaining rather than testing, one may, in the deductive-nomological fashion, subdivide the observational record (iii). There is, for one, a statement of initial conditions (here roughly, “in the experimental setting, light travels through both environments”), and for another, a statement of the experimental result (here roughly, “in the experimental setting, light is *less fast* in water than in the air”). However, if the purpose is only to explore the underdetermination thesis, it is acceptable to collect these statements into a large conjunction, and this is indeed the format of the received discussions of Duhem and Quine.¹² Notice further that our classification cuts across logical distinctions made in terms of levels of generality. By nature, (iii) is made out of singular statements, but both (i) and (ii) can contain singular statements, such as those assigning a value to a physical constant, beside universal statements of either putative laws or *de facto* regularities.

Abstracting from his physical applications, we can state Duhem's underdetermination thesis as follows:

(a) *When a theory is tested against some piece of evidence, and this results in a contradiction between the test prediction and the observational record, there is no compelling way of deciding that some of the primary hypotheses, rather than some of the auxiliary hypotheses or some of the statements of the observational record, are false.*

This roughly corresponds to what has been called “the weak form of the Duhem-Quine thesis”, in contrast with a “strong form”, which in the present framework would read as follows.

(a') *When a theory is tested against some piece of evidence, and this results in a contradiction between the test prediction and the observational record, it is possible to change the auxiliary hypotheses in such a way that they entail the observation record in the presence of unchanged primary hypotheses.*

The so-called strong form cannot be found anywhere in *TP*, but is evocative of a famous claim in *FLPV*: “Any statement whatever can be held true come what may, if we make drastic enough adjustments elsewhere in the system” (1953, p. 43). It is indeed in connection with Quine alone that it has normally been addressed.¹³

These formulations are more or less informative, depending on what is put in the essential words “compelling” in (a) and “possible” in (a’). It is no great news that the truth-tables of propositional logic do not decide where the falsity lies in a false conjunction; so if one actually means “compelling in the sense of standard logic”, there is something painfully obvious about (a). By the same token, “possible” must be restricted so as to preserve (a’) from the crude move that consists in replacing the auxiliary assumptions with a statement to the effect that the primary hypotheses imply the observational record. Nevertheless, the two variants do not react similarly when trite possibilities are allowed. For even if one takes “compelling” in the standard logical sense, one learns something useful from (a), which is *the range of statements to which the refutation decision can be applied*. Duhem identifies three possible sources of falsity, but another philosopher, even another holist, might have concluded differently. In other words, what matters is not so much the thesis itself as the kind of epistemological holism from which it derives. By contrast, nothing remains of (a’) if one does not restrict the meaning of “possible” appropriately. As it happens, those philosophers – prominently Grünbaum – who were prepared to give some mileage to the “strong Duhem-Quine thesis” have been unable to clarify what would in this context distinguish an acceptable from an unacceptable change in the auxiliary hypotheses.¹⁴ If one were to ask empirical scientists to tell on particular examples what this difference is, counterexamples to (a’) would presumably abound. The underdetermination thesis understood this way seems to leave no middle ground between its trivial interpretation and the claim that it is false.¹⁵

Both because of the conceptual difficulties surrounding (a’) and the fact that Duhem never considered it, we will take (a) to express the underdetermination thesis. Even this apparently unassuming formulation raises a number of problems, which we will now briefly discuss.

2.2 *On Interpreting Duhem’s Underdetermination Thesis*

One problem has already been touched on. If the choice of a logic, *in the sense of some prescriptive formalism*, is left open, (a) makes a very assertive claim. Thus, beyond what can obviously be said for the propositional logic case, it would deny that the Bayesian conditional probability apparatus, the theory of statistical inference, and – for what can be made out of it – the Popperian calculus of “corroboration”, can decide which hypothesis is responsible for a refutation. (With a Bayesian or statistical interpretation, “contradiction” would have to be replaced in (a) by a more appropriate word.) An even stronger claim is that there is no *rational* selection of the statements refuted in the conjunction, and formulations like this have indeed circulated in connection with the “Duhem-Quine thesis”.¹⁶ However, the specialized literature has usually stuck to the reading of (a) in terms of propositional logic alone, and we think that it is right on this score. For there is a natural division of issues here – first, what does the range of underdetermination, in the straightforward propositional sense, exactly consist of, and second, can this range be narrowed

down by moving to some other, more powerful, prescriptive formalism? For clarity, the last question should be treated not under the heading of the underdetermination thesis itself, but as an answer to the selection problem it raises.¹⁷

Second, formulation (a) involves the refutationist analysis of empirical tests, but what about its *verificationist* variant? The suggestion has occasionally been made that the underdetermination thesis should apply to refutation and verification symmetrically. The variant would say that there is no compelling way, in the propositional logic sense, of deciding which hypotheses are verified when the observational record entails the test prediction. Such an underdetermination thesis can only be of interest to those who think that statements can be verified by their consequences – not a large crowd in past and present philosophy of science, and Duhem is not a member of it. The quest for a counterpart to the refutationist variant makes more obvious sense if probabilistic connections replace propositional inferences in the analysis of empirical tests, but this also means that verification gives way to probabilistic confirmation, and an altogether different range of problems arises.¹⁸

Third, there will typically be several primary hypotheses to consider when (a) is applied. Duhem made this point forcefully in connection with Foucault's experiment; in his words, this physicist's concern was the corpuscular *system* ("système de l'émission"), and not the corpuscular *hypothesis* ("hypothèse de l'émission", *TP*, II, Chap. VI, §II, p. 283; English trans., p. 187). More will be said on this score after the next, related comment.

Fourth, Duhem's examples also make it clear that auxiliary hypotheses are multifarious; they are not only singular statements or lower-level generalities, as the particular test requires, but also law-like statements that may compare with the primary hypotheses in terms of depth and coverage. Thus, beside highly specific assumptions, Wiener had to assume the standard laws of optics when he set up his experimental device; in this case, the auxiliary hypothesis was no less substantial than Neumann's hypothesis itself. This suggests that the distinction between groups (i) and (ii) should be drawn in *pragmatic* and *epistemic* terms rather than in logical or semantic ones. Roughly speaking, (i) collects those theoretical statements which motivate the empirical inquiry and are currently regarded as problematic, while those in (ii) are introduced on grounds of convenience and at least in the initial intent, not taken to be problematic. Observe that group (i) would be inflated dramatically if it were assumed to include a theory of the scientist's perception. The standard laws of optics, as well as many other physical and non-physical laws, would then be involved quasi-automatically in the range of underdetermination. As it appears from *TP*, Duhem resists this slippery line. His comments on Wiener relate to the optical theory underlying the physical device, not that underlying the physicist's unaided visual perception. Admittedly, some passages of *TP* extend holism to the point of fusing the sets (i) and (ii), and equating them with the whole of physics, but even there, Duhem does not seem to endorse the strong argument we are now envisaging.¹⁹

Previous discussions of the underdetermination thesis have relied on a separation different from the present one between *one* privileged hypothesis *H*, the various auxiliary hypotheses *A*, and the observational record *O*. Underlying this scheme is

the view that an empirical test is directed towards a specified hypothesis. All other hypotheses in (i), even if they belong to the same scientific theory, are pooled with the auxiliary assumptions of (ii). The pool contains all the statements that the scientist can logically blame for the refutation while having no pragmatic or epistemic interest in doing so, since he wishes to relate O to H , not to any of the A . This schematic formalism is sufficient for some argumentative purposes, like the comparative assessment of the “strong” and “weak Duhem-Quine thesis”, but misleading as a rendering of Duhem’s thesis in general. It does not fit with Foucault’s case, in which a complex theory was subjected to the test. If it were argued that the formalism takes this case into account because H can stand for a whole theory, the objection would become that this formalism is too loose, because it hides the difference between Foucault’s case and that in which H effectively refers to a single hypothesis. There is another significant case to consider, which we will indeed illustrate with EUT: the scientist aims at checking individual components, but in actual practice, tests only large chunks of the theory, because some components turn out to be inseparable. The diversity of empirical situations suggests replacing the standard formalism by the slightly richer representation adopted here.

2.3 *Duhem’s Thesis Against Crucial Experiments*

Before Quine advertised Duhem for his epistemological holism, he had been noted for his unusual dismissal of *crucial experiments*. The argument is carried out in a condensed paragraph of *TP* (II, Chap. VI, §III) significantly entitled “A ‘crucial experiment’ is impossible in physics” and makes use of the theory of light again. This example is right to the point, because Foucault’s experiment was implemented after a preexisting scheme by Fresnel, who had invented it with the view of finally resolving the conflict between corpuscular and undulatory hypotheses. It would appear that Fresnel conceived of the experiment, if it could be carried out at all, as being both decisive for one hypothesis and against the other. This is the strong sense commonly given to crucial experiments in empiricist philosophies of science, and also the sense that Duhem is concerned with dismissing, but his argument turns out to be relevant to a weaker notion as well. Let us define a *simply refuting* crucial experiment as an experiment which decides against – in the sense of refuting – one of the two theories without necessarily deciding for – in the sense of either verifying or confirming – the other. This restricted meaning is due to Popper, who, from the *Logic of Scientific Discovery* onwards, has repeatedly claimed that it was both important for science and immune to Duhem’s criticisms.²⁰ With this internal distinction at hand, we state Duhem’s second thesis:

(b) *There are no crucial experiments in empirical science, even simply refuting ones.*

When Duhem begins to deny the existence of crucial experiments, he seems only to expand on his already explained underdetermination thesis, but this is deceptive appearance. The two claims are clearly distinct, and it is implausible that Duhem

identifies them. In fact, he derives (b) from the same arguments he had used for (a), relying on the continued analysis of the same example. In Fresnel's experimental scheme, a light ray is sent through water, another through the air, one above the other, and a rotating mirror reflects each, producing two spots of light moving on a screen. Supposedly, the location of the spots indicates which of the three following statements holds true: light is as quick in one medium as the other, quicker in water, as the corpuscular hypothesis lets one to expect, or quicker in the air, as the undulatory hypothesis would rather have. Now, the arguments for thesis (a) could be repeated with a similar damaging effect on (b): the predictions on respective speeds follow from whole systems of hypotheses, the postulated connection between these test predictions and the observation of moving spots depends on auxiliary hypotheses, etc. More briefly, if (a) holds, (b) follows, since a crucial experiment is at least decisive against one of the two hypotheses, and (a) says that hypotheses are never decisively refuted. It is important to state the argument in this abstract fashion, which is not Duhem's, because it then becomes clear that it hits the notion of a simply refuting experiment no less than the stronger, more familiar notion.

Duhem has another – by now classic – argument against crucial experiments, which is entirely different from the use of either (a) or the reasons for (a). In a nutshell (and again in abstract terminology), when scientific hypotheses conflict with each other, they are not *contradictory*, but simply *contrary*, with each other; that is, they cannot be true together, but can be false together. *TP* makes this point when denying that two hypotheses alone can ever exhaust the physical possibilities (“do two hypotheses in physics ever constitute such a strict dilemma?”, p. 288; English trans., p. 190), and what amounts to essentially the same, when claiming that *reductio ad absurdum* is irrelevant to physics (“unlike the reduction to absurdity employed by geometers, experimental contradiction does not have the power to transform a physical hypothesis into a indisputable truth”, *ibid.*).²¹ Duhem also objects to inductive elimination by arguing that it would work only if the stock of conceivable physical laws were finite, which is impossible. It is easy to see that this is but an iterative generalization of the present argument against crucial experiments.

Actually, this whole line of reasoning seems to be dubious, since it represents the empirical test as a duel between putative laws and observations, which (a) precisely denies to be the correct description. As some have rightly pointed out,²² Duhem can use the argument on *reductio ad absurdum* only as a reinforcement of his initial position: *even if, implausibly, tests could be construed as duels*, it would not follow that crucial experiments exist. This may be helpful to add against hard-liner empiricists, but will normally need not mentioning. If the argument has limited dialectical use, this is because it is weaker than the main one through (a). It denies the verificatory component of crucial experiments, which is enough for the conclusion that they do not exist in the strong sense, but simply refuting crucial experiments escape its strictures.

By this logical observation, one may hope to rationalize Popper's problematic claim that “Duhem in his famous criticism of crucial experiments succeeds in showing that crucial experiments can never establish a theory. He fails to show that they cannot refute it” (1963–1972, p. 112, n. 26). It would seem as if Popper approved

of Duhem's exclusion of *reductio ad absurdum*, with its pleasant antiverificationist consequence, but rejected his argument through (a), which means a challenge to refutationism. If this is indeed Popper's move here, it is distressing, because he generally accepts the underdetermination thesis, at least in some form. As early as *The Logic of Scientific Discovery* (henceforth *LSD*), he envisaged several reasons why empirical refutations may not be compelling, in particular mentioning two Duhemian sources, i.e., the presence of auxiliary hypotheses and the equivocality of observations; see the passage on "conventionalist strategies" (1935–1972, §9 and §19–20). To avoid an inconsistency, there is only one interpretation left, i.e., that Popper uses the idea of a refuting crucial experiments loosely. He just means a testing experiment that strongly supports the conclusion that a certain hypothesis – or failing this, a certain compound of hypotheses – is false. But in this diminished sense, Duhem fully accepts that refuting crucial experiments exist, as we now proceed to explain, and Popper has no reason for claiming a serious disagreement with him.

2.4 *Duhem's Answer to the Underdetermination Problem*

We have been careful to distinguish between Duhem's underdetermination *thesis*, which is a bare statement of the possibilities for refutation, and Duhem's corresponding *problem*, which is to turn one of these possibilities into a final refutation. As we have reconstructed it, the thesis relies on propositional logic alone and escapes triviality only because of the range of statements that it specifies. With this construal, it does not entail that there is no rational way out of the underdetermination, let alone that there is no logical way out, if logic is taken beyond the propositional realm – these conclusions depend on how the problem is appreciated. Now, Duhem's answer to the latter is far from being clear-cut, but this is in part of necessity, because it is a pragmatic, and as we will argue, a historical answer.

When certain consequences of the theory are struck by experimental contradiction, we learn that this theory should be modified but we are not told by the experiment what must be changed. It leaves to the physicist the task of finding out the weak spot that impairs the whole system. No absolute principle directs this inquiry, which different physicists may conduct in very different ways without having the right to accuse one another of illogicality. For instance, one [physicist] may be obliged to safeguard certain fundamental hypotheses while he tries to reestablish harmony between the consequences of the theory and the facts by complicating the schematism in which these hypotheses are applied by invoking various causes of error, and by multiplying corrections. The next physicist, disdainful of these complicated artificial procedures, may decide to change some one of the essential assumptions supporting the entire system... Each is *logically* permitted to declare himself content with the work that he has accomplished.

That does not mean that we cannot very properly prefer the work of one of the two to that of the other. Pure logic is not the only rule for our judgments; certain opinions which do not fall under the hammer of the principle of contradiction are in any case perfectly unreasonable. These motives which do not proceed from logic and yet direct our choices... constitute

what is appropriately called good sense (“ce qu’on appelle proprement le bon sens”, *TP*, II, VI, §X, p. 329–330; English trans. p. 216–217).

When comparing the conservative and the radical physicists of this passage, an everyday observer might be repelled by the “obstinacy” of the first, or to the contrary, by the “haste” of the second. Duhem does not say which of the two opposite attitudes carries more weight, and does not even suggest that one of them must prevail over the other; there may be a stalemate, as it were. Our interpretation of this passage emphasizes not commonsense, an elusive concept, but the *pattern over time* involved in the psychological description. The conservative physicist becomes unbearable only when he repairs the existing theory *over and over*, and the radical is unpalatable only when he strikes at it *too early*. Here is how Duhem redescribes the conflict in the theory of light:

Biot by a continual bestowal of corrections and accessory hypotheses maintained the emissionist doctrine in optics, while Fresnel opposed this doctrine constantly with new experiments favoring the wave theory.

In any event this state of indecision does not last forever... After Foucault’s experiment had shown that light traveled faster in air than in water, Biot gave up supporting the emission hypothesis... By resisting... for a longer time... Biot would have been *lacking in good sense* (*TP*, II, VI, §X, p. 331; English trans., p. 218).

Putting aside the allusion to commonsense, we find our time-based interpretation reinforced by this quote. Foucault’s experiment came after the same situation had been repeated – one group multiplying reinterpretations, while the other brought out many new facts and hypotheses. The experiment draws its significance from its place in an history. It is *critical* like the turning point of an illness in traditional medicine, a sense which is easy to contrast with the cruciality denied by (b).

More abstractly, we identify Duhem’s answer to the underdetermination problem with the following claims. (i) Sometimes in the history of science, this problem received a satisfactory solution. (ii) These solutions typically arose in a conflictual environment, after a lengthy exchange of arguments in the face of renewed evidence, with the arguments weighing more for one camp from some moment onwards. (iii) These solutions ended up in the qualified rejection of a particular hypothesis, despite the fact that the discussion had involved many other statements and not all of them were checked properly. If Duhem’s final position does include (iii), it goes beyond what is sometimes said of him, i.e., that he took refutation to be capable to hit complex theories and stopped at this conclusion. When Duhem moves from logic to pragmatics and history, we understand him as suggesting that what was eventually refuted in the 19th century theory of light was the corpuscular theorists’ *hypothesis*, i.e., that light is made out of particles, and not simply the *system* in which this hypothesis was encapsulated.

The informality of this resolution stands in sharp contrast with the more definite proposals made by later philosophers of science, especially the Bayesians and the falsificationists. The latter make for an easy comparison because of the refutationist ground they share with Duhem. Popper’s own answers to the underdetermination problem follow two rather different lines.²³ The first amounts to claiming that the underdetermination problem arises only for those who search for certainty, like the

verificationists. Once we are reconciled with the view that all our scientific hypotheses are guesses anyhow, we stop worrying whether or not we blame the wrong subset of premisses for the refutation. This seems very much like throwing the baby with the bathwater. When Popper takes the second line, he must believe that the underdetermination problem matters after all, since he tries to solve it. Essentially, he recommends that the scientist reformulate his premisses so as to make them logically independent of each other, and once he has done so, that he devise a test for each in turn. Unlike the previous one, this is a very substantial suggestion, and indeed, one which scientists have naturally abided by. We will see the rôle it played in the decision theorists' assignment of responsibilities between the different components of VNM theory when the latter became submersed with counterexamples. However, this procedure of *independent reformulation and testing* is clearly not available everywhere: it may fail at either of the two stages, i.e., the logical decomposition of premisses may be lacking in independence, and there is no guarantee anyhow that premisses and tests can be associated in an one-to-one way (we will see that this happened in the VNM case). Moreover, the method is helpful only to spot a falsehood *in the primary hypotheses*, and it works mostly for theories in which these hypotheses have already been axiomatized (VNM theory being one such case). It is unclear how to apply it, except very informally, outside this realm of applications.

About these other sources, Popper has something important to add:

As regards auxiliary hypotheses we propose to lay down the rule that only those are acceptable whose introduction does not diminish the degree of falsifiability or testability of the system in question, but, on the contrary, increases it (1935–1972, § 20, pp. 82–83).

This is excerpted from the already mentioned passage of *LSD* about “conventionalist strategies”, i.e., those moves which deflect the refutation from the target hypothesis when the empirical test is negative. It is in this indirect way that Popper faced the underdetermination problem in his early work. We may generalize his criterion for changes in the auxiliary hypotheses to *all* logically possible modifications of the Duhemian compound. To apportion refutation between the candidate statements, first compare the systems following from each apportionment in terms of their overall refutability. Then, declare an apportionment to be *acceptable* if it results in a system that is more refutable than the initial one and it is not itself refuted empirically. Thus adapted, the rule delivers a crisp necessary condition for an answer to the underdetermination problem. Comparison with Popper's later work, especially *Conjectures and Refutations*, suggests that it is well representative of his long-run thinking on this problem. It agrees with his general claim that *a necessary condition for scientific progress is that the new theory will have more testable content than the current one*.²⁴ We may add that a *best* apportionment within the acceptability class is one which results into a maximally refutable system, in the sense that no other acceptable apportionment generates a more refutable system. Again, this agrees with what Popper argues about scientific progress.

The straightforward objection to these suggestions is that they are too strict. To illustrate by an example loosely related to EUT, it would not permit replacing the false regularity:

- (1) “All individuals, in all their dealings, are risk averse”. by, say, the following one, which is not so obviously false:
- (2) “All individuals with a low income, in all their dealings except those concerned with gambling with small stakes, are risk averse”.

Despite the sharp restriction of scope, hence the decrease in refutability, it is hard to justify that generality (2) should be excluded from consideration. On further examination, it may turn out to involve an optimal mix of content and approximate truth among the variants of (1). To define an acceptable apportionment in the way just said will block (2) from competing for the status of a best revision. We elaborate here on an objection made by Lakatos (1970, p. 182) to the effect that Popper disallows restrictions in the antecedent of well-established laws that serve to absorb anomalies. This is too specific a way of putting the objection, because the need for introducing restrictions of false generalities can arise from many sources. If it is absurd to reject them off hand, this is simply because they may be the only true generalities that are to be found. Lakatos is still too much of a Popperian in order to accept the possibility that empirical sciences can make genuine advances *while losing in overall refutability*, and he eventually bans it from the definition of progress that regulates his “methodology of scientific research programmes”.²⁵

Although himself a refutationist, Duhem gives no sign of having ever envisaged Popper’s and Lakatos’s ban. If our EUT case study carries any normative force, it will suggest that Duhem was right, and the two falsificationists wrong. For decision theorists groped their way from the refuted VNM theory towards alternatives that involved losses of content. This is one of the several challenges posed by their work, which we now proceed to explain.

3 The Formative Stage of Von Neumann–Morgenstern Theory

3.1 *Von Neumann and Morgenstern’s Contribution to Expected Utility Theory*

This case study is concerned not with expected utility (EU) theory as a whole, but only with its more elementary part, which is referred to as *von Neumann-Morgenstern (VNM) theory*. The standard exposition consists of a small body of axioms – usually three and at most five – followed by a few salient consequences, the most important one being the VNM *representation theorem*. According to the latter, if an individual’s preference relation is defined on risky prospects involving given probabilities (“lotteries”) and conforms to the axioms, this preference relation gives rise to a numerical function having the *expected utility form*, i.e., the form of a sum of the utility values of outcomes weighted by the probabilities of these outcomes. The function *represents* the individual’s preference relation in the sense that, for any pair of lotteries, it gives a higher value to the one which is strictly preferred to the other (and it gives equal values to indifferent lotteries). This is the *existence* part of

the VNM representation theorem. The *uniqueness* part says that the EU form of the representing function identifies it uniquely up to positive linear transformations. Exact statements will be given in the next subsection.

VNM theory is a paradigmatic example for contemporary choice theories. All these theories rely on a sharp distinction between information relative to (an individual's) *preference* and information relative to (this individual's) *utility*. Only the former is the substantial concept; the latter is but a numerical index representing, in the technical sense explained, the agent's preferences or choices (for current purposes, we may follow the questionable habit taken in the field of disregarding the significant difference between these two notions). This conceptual hierarchy sets a target for an axiomatization of the choice theory. The symbols appearing in the axioms should concern either the objects of preference or the preference relation, while the symbols of utility should be reserved for the logical consequences of the system, such as the representation theorem. This very clear structure became pervasive in the 1950s under the influence of von Neumann and Morgenstern (1944–1947), the eponymic founders of the theory we are concerned with.²⁶

VNM theory takes as given the probabilities from which it derives the EU representation, so its limited task is to derive a utility function on the outcomes and combine this output in the EU way with the preexisting probabilities. A primitive concept in the VNM axiom system, the *lottery* is a mapping associating probability p_1 with outcome x_1 , probability p_2 with outcome x_2 , and so on; abstractly, it is a probability measure p on the set of outcomes X . A more sophisticated theory, Savage's (1954–1972), *derives* the probability measure p at the same time as the u function, and shows it to be unique given the axiomatic system. The latter replaces lotteries by other objects of preference, the *acts*, which involve no prior quantitative information. Many, especially in economics, contrast the two branches of expected utility theory as *objective* and *subjective*, respectively, but from the philosopher's point of view, these are misleading expressions. Savage's endogeneous probability measure can be interpreted as being subjective, in the sense that they express the agent's degrees of belief, but there is nothing in von Neumann and Morgenstern to warrant an objective interpretation of the lotteries, even in a less specialized sense of "objective" than frequentism. There is no more to be said in general of these objects of preference than that they are *given*. Admittedly, they often refer to actual lottery tickets with a physical device underlying them, such as those sold by the National Lottery, but they may also represent combinatorial possibilities in a game of chance, or an agent's subjective probabilities, however this agent arrived at these estimates. In brief, VNM theory is too terse to allow for a single interpretation, a warning that also applies to the hypotheses eventually proposed to replace it.

In *Theory of Games and Economic Behaviour* (1944–1947), von Neumann and Morgenstern gave the impetus to the theory labelled after them, but their technical treatment differs from that which came to be received afterwards. Paradoxically, the book puts in place the style of axiomatization now received in choice theory without fully abiding by it. Chap. I,3 of the 1944 edition and the 1947 Appendix developing this chapter are expressly devoted to "the axiomatic treatment of utility". Contrary to what this expression would now suggest, von Neumann and Morgenstern do not

take the preference relation on lotteries as a primitive concept. Instead, they start from comparisons bearing on equivalence classes of lotteries, as are defined by the indifference relation. Their work consists in axiomatizing comparisons made between these peculiar objects, and from these comparisons, demonstrating the existence and relative uniqueness of the EU representation.

The previous discussion may seem to be abstruse, but it goes a long way to account for the strange start made by VNM theory. The truly substantial axiom of this theory – that which was to matter dramatically to both its empirical testing and normative assessment – is the so-called *von Neumann-Morgenstern independence condition*. But remarkably, it came to be identified only after the founders, and this is because they had hidden it behind their technical choice of objects of comparisons. This slip was pointed out by Samuelson and Malinvaud rather belatedly – at a conference held in Paris in 1952, which was to have other lasting consequences that we will soon review.²⁷

At that time, a reasonably explicit set of axioms, due to Marschak (1950), was already available. Then came the entirely transparent ones of Friedman and Savage (1952), Samuelson (1952a, b), Herstein and Milnor (1953), Luce and Raiffa (1957), and still others. Each of these works obeys the pattern of today's choice theory by moving from the preference relation on lotteries to the EU representation, and each includes a relevant version of the – ironically called – VNM independence condition.

3.2 *An Axiom System for VNM Theory*

We will discuss the axiomatic work in terms of what may be the easiest system of all, i.e., Friedman and Savage's (1952). Its axioms strongly overlap in logical content, which is a defect from the logician's perspective, but makes it more tractable than more independent systems like Herstein and Milnor's (1953). A *lottery* is defined to be a probability measure p on a set X , called the set of *outcomes*. Assuming finiteness for convenience, we can write

$$X = \{x_1, \dots, x_i, \dots, x_n\} \text{ and } p = (p_1, \dots, p_i, \dots, p_n) \text{ with } p_i \geq 0, \sum p_i = 1$$

(that is to say, probability measures are simply probability vectors). A *weak preference* is a binary relation R on the set L of all such vectors, with " $p R q$ " reading as " p is strictly preferred or is indifferent to q ". The system consists of three axioms on R and L :

- (A1) (Ordering) R is an ordering (i.e., it is reflexive, transitive, complete) over L .
- (A2) (Continuity) R is continuous.²⁸
- (A3) (VNM Independence) For all p, q, r in L , and all numbers a between 0 and 1 (0 excluded), $ap + (1-a)r R ap + (1-a)r$ if and only if $p R q$.

The VNM representation theorem follows. Formally: if (P1), (P2), and (P3) hold, then

(Existence) there exists a numerical function u on X such that

(*) $p R q$ if and only if $\sum p i u(x_i) \geq \sum q i u(x_i)$,

and

(Uniqueness) the functions u' on X satisfying the same property (*) as u are exactly the functions $u' = \alpha u + \beta$, where α is a positive number and β any number.

Both axioms (A2) and (A3) involve convex combinations of elements in L that one would like to interpret as *compound lotteries*, i.e., lotteries the outcomes of which are lotteries, not final outcomes. However, mathematically, a convex combination of probability vectors is another probability vector, which means that the chosen formalism cannot express the distinction between compound and simple (non-compound) lotteries. Take for example:

$p = (5 \text{ M FF with prob } 0.10, 1 \text{ M FF with prob } 0.89, 0 \text{ FF with prob } 0.01)$, and

$p' = (1 \text{ M FF with prob } 0.89, r \text{ with prob } 0.11)$, where $r = (5 \text{ M FF with prob } 0.10, 0 \text{ FF with prob } 0.01)$.

Once the relevant multiplications are made, p and p' become the same 3-dimensional probability vectors. So the formalism leaves no choice but to identify the corresponding lotteries, which is objectionable because the agent may react differently to them. Luce and Raiffa (1957, p. 26) address this problem by redefining lotteries as lists of *prizes* and corresponding probability numbers, each prize being a lottery of lesser complexity, which may be an outcome as a particular case. This formalism allows for much flexibility in the description of the agent's preferential attitudes. Luce and Raiffa do not pursue the topic, and are content with imposing the axiom of *reduction of compound lotteries*, to the effect that the agent reacts in the same way, whatever the stage of reduction of the lotteries presented to him. This leads back to the standard assumption, but makes it explicit, instead of burying it in the definition of a mathematical primitive. The higher degree of explicitness facilitates the interpretation and testing of the overall system, so we will take Luce and Raiffa's axiom to be a component part of VNM theory, even if we do not state it formally.²⁹

Following this empirical line, we list a few situations in which the supplementary axiom would be violated. (a) The compound lottery stands for a physical device that does not satisfy the probabilistic independence of the successive stages, so that the associated simple lottery cannot be computed without more probabilistic information being given. (b) In the probabilistic independence case, the individual does not know how to make the calculations. (c) The individual feels a specific interest in participating in a many-stage game rather than in playing just once. To adapt an example from Ellsberg (1954, pp. 543–544), a player in a sequential poker game might choose to stay in this game even though he would not accept the onestage game with the calculated multiplicative probabilities. This gambler likes being involved, and this is why he stays in despite a hand which he would otherwise judge to be unpromising. The example falls under the heading of *the pleasure of the game objection*, which VNM theorists have acknowledged unproblematically. By contrast, they pay little or no attention to objections (a) and (b), perhaps because many concrete examples of lotteries do satisfy the probabilistic independence assumption denied in (a), and because (b) is just another intractable example of bounded cognitive ability. Whatever the differences of theoretical attitude vis-à-vis (a), (b)

and (c), many would regard failures of the reduction of compound lotteries as constructive suggestions on how to restrict the domain of application of VNM theory appropriately.

3.3 Ordering and Continuity

By a well-known result in choice theory, if a preference ordering on some set of alternatives is continuous in a suitable technical sense, there exists a continuous numerical function on this set representing the preference, and it is unique up to monotonically increasing transformations. Comparing this statement with (A1) and (A2), on the one hand, and the VNM representation theorem, on the other, one can see what the contribution of (A3) consists of. Essentially, (A1) and (A2) deliver a representation V of R on L that is left unspecified, except for the weak properties of being continuous and equivalent only to its monotonic transforms (like V^2 or $\text{Log } V$, and unlike $-V$ or $1/V$). Adding (A3) has the effect of restricting both the functional shape of V (by conforming it to the EU formula, or in mathematical terms, making it linear in the probabilities) and the class of its admissible transformations (V is then equivalent only to monotonic linear transforms). That it took time to isolate (A3) can partly be explained by the fact that early VNM theorists were not yet accustomed to the mechanics of representation theorems.³⁰

How are the two axioms to be assessed? Following a widespread view in both choice theory and microeconomics, (A1) states a minimal requirement of rationality, thus being unobjectionable normatively, and it is also a roughly plausible empirical generality. Expected utility theorists have most of the time refrained from questioning it, and we will not discuss it in this paper. Concerning (A2), the predominant view is that it is a technical condition without any clear empirical counterpart serving as either evidence or counterevidence. Interestingly, however, not every VNM theorist has followed this non-committal line. Marschak (1950, p. 117) demonstrates that in the presence of the ordering property and a relevant part of VNM independence, his version of continuity leads to the following:

(C) For all p, q and a strictly between 0 and 1, if $p P q$, then $p P ap + (1-a)q P q$. (P stands for the *strict* preference relation that is induced by R .)

This is tantamount to saying that an individual faced with a compound lottery having two outcomes will not value this lottery more than the outcome he values more, and not value it less than the outcome he values less. Thus, (C) sounds like an acceptable conclusion, both normatively and empirically. But Marschak puts forward an intriguing counterexample that is worth reconsidering in detail.

A typical mountain-climber dislikes being alive for sure ($= p$) and incurring death for sure ($= q$), ranking the former of these two (here degenerate) lotteries above the latter, whereas he likes risking his life with a 95% chance of surviving the climb ($= ap + (1-a)q$). That is to say, he strictly prefers a compound lottery to *both* of its components, and not only to the less preferred one, thus violating

(C). Marschak's (1950, pp. 138–140) comment stresses that the same “love of danger” can also underlie activities – like the funding of scientific research or the undertaking of geographic expeditions – that are standardly classified as being “economic”. The proper distinction, Marschak goes on, is between *two different ways of taking chances*, and it will typically cut across the same spheres of human activities. So far, so good; but then, surprisingly, Marschak does not try to identify the two ways in terms of relevant psychological, behavioural, or situational properties. Instead, he discusses the sense in which VNM theory can account for chance-taking action.

Given a suitable set of outcomes X (the easiest example being a set of monetary amounts), the theory defines an individual to be a *risk-lover* if the u function delivered by the VNM representation theorem is convex, and a *risk-avorter* if it is concave. The concept of risk-love is the single tool made available by VNM theory to explain gambling, lack of insurance, speculation, and the like, and it works well on many activities, but it also fails on some, and the mountain-climber is a sign-post for these exceptions. The turn in Marschak's discussion suggests that it must have been difficult, even perhaps impossible, for him to assign a domain to VNM theory *from the outside*. He resorts to the theory *itself* to say where this theory applies. His argument is not vacuous because he makes two relevant points: firstly, that exceptions of the rock-climber sort are essentially unmanageable for the theorist (a violation of (C) excludes the existence of a utility representation, whether VNM or not), and secondly, that they are best regarded as being irrational.

We have gone through this intriguing discussion because it reveals salient features of VNM theory – and perhaps even of choice theories – at large. First, it confirms that the theory was accompanied with recognized counterexamples from its very beginning, these counterexamples being regarded as not particularly threatening. The tendency was to view them as suggesting domain restrictions rather than stating refutations or even anomalies. For historical reasons that we cannot develop here, *the pleasure of gambling* functioned like a convenient buffer zone for various problematic cases.³¹ We have seen this concept at work in Ellsberg's discussion of Luce and Raiffa's compound lottery axiom. Marschak inherits its defensive use and attempts to refine it by separating those cases of the gambler's pleasure which VNM theory can redeem and those which it cannot, and here we record a second noteworthy point. Instead of invoking the nature of the chance-taking actions, the theorist directs attention to the *theoretical tractability* and the *rationality*, as assessed by himself, of these actions. From now on, the evaluation of empirical performance of VNM theory will be mixed with an assessment of these qualities, despite what may seem to be an irrelevant association. Third, the theorist does not have a preexisting yardstick of rationality, but argues about it in a circle of reinforcing considerations. Daring mountain-climbers suggest an initial diagnosis of irrationality that is being confirmed by the technical observation that their preferences clash with the axioms, and in the particular instance, with any smooth mathematical description. The tractability argument here comes to the rescue of the rationality attribution.

3.4 Early Theoretical Attitudes Towards VNM Independence

With the benefit of hindsight, one would have expected the very substantial independence condition to become the focus of theoretical attention once it was recognized, but this was very far from being the case. Those who filled the gap in von Neumann and Morgenstern's axiomatic system believed that it was an unproblematic addition to make. The first to formalize independence, Marschak wrote that it was "very weak" (1950, p. 111) and spent less time on it than on continuity. At an already late stage, Ellsberg (1954) offers a stunning example of this general neglect. Having recalled the entrance of (A3) into the VNM axiom system, he claimed that "it seems rather hard to justify this emphasis since the axiom seems indubitably the most plausible of the lot" (1954, p. 544), and, like Marschak, focussed his discussion, both normative and empirical, on the other conditions.³²

At least Friedman and Savage (1952, pp. 468–469) and Samuelson (1952a, pp. 133–134) felt the need for a justification of (A3). The former claim that (A3) follows from an elementary principle of choice theory, the *dominance principle*, which in the particular context would go like this:

If an agent weakly prefers lottery p to lottery q in case event E occurs, and weakly prefers lottery r to lottery s in case the complementary event E^c occurs, then this agent should weakly prefer the prospect giving p in case of E and r in case of E^c to the prospect giving q in case of E and s in case of E^c .

Virtually no decision theorist objects to the reasonableness of this principle, so it would buttress (A3) if this condition did logically follow from it, but this is not the case. The dominance principle relates to *prospects*, which are defined vis-à-vis states or events, not lotteries, which are stated in terms of probabilities, and this turns out to make a world of difference. The dominance principle can be reformulated to fit VNM theory, but when this is done, it turns out that (A3) implies it, and not the converse.³³

As for Samuelson, who boasted having become a convert to VNM independence once he knew how to formulate it, his justification amounts to the following. Consider two compound lotteries $l' = aq + (1-a)r$ and $l = ap + (1-a)r$, and let us ask how the agent would assess them at the *interim* stage, that is, after these lotteries have been drawn, resulting in either p for the first and q for the second, or r for both of them, and before these lower-level lotteries are drawn. At this stage, the choice between l and l' becomes one between p and q , since there is nothing left to choose in the case where r has been drawn. Taking for granted that the *ex ante* preference between l and l' should conform to what the interim stage will be, Samuelson concludes that (A3) is warranted. Contrary to Friedman and Savage's, this argument initiates a conceptually fruitful line. It foreshadows the later *dynamic analysis of EUT*, in which many today see the strongest possible normative arguments that can be made for VNM independence (or the related condition in Savage).³⁴ However, as this more recent work makes clear, more than one assumption must be made for Samuelson's argument to be logically complete, each being normatively questionable. The brief and authoritative comment of the 1950s does not do justice to the complexity of the logical and conceptual issues it raises.

3.5 A Holistic Argument for VNM Theory

There was another course to take for those early writers, mostly economists, who were trying to protect the budding VNM theory against premature dismissal. Instead of discussing individual conditions or predictions in the face of partial evidence, they could, despite the still limited empirical record, inquire about the performance of the theory as a whole. Equally, instead of trying to assess it in isolation, they could place it against the background of earlier treatments of choices under risk in economics. This double shift towards holism, with essentially soothing conclusions, can be found in the first paper by Friedman and Savage (1948), which forms an interesting contrast with the second.

The two writers go back to Marshall's *Principles* (1890–1921), a treatise which was still influential in Anglo-American quarters at the time. The passages of this book which discuss gambling dash any hope of handling any behaviour of this kind, even of an economically relevant one, with the tools of economic theory.³⁵ Marshall's argument is already couched in terms of expected utility calculations, which he performs under the assumption that the u function is concave – a reflection of his intangible principle that marginal utility is decreasing. With this assumption, any form of gambling involves a loss, insurance always involves a gain, and it is therefore irrational to gamble and not to insure. The final implication is that if economics is to remain the science of rational conduct, there is no place in it for an analysis of gambling and – more importantly, of course – of the broadly similar economic activities in industry, commerce and finance. This appears to be a severe self-limitation of the discipline.

Friedman and Savage (1948, pp. 280–281) contrast von Neumann and Morgenstern's work with this discouraging Marshallian line, which they correctly take to be typical of earlier economics. They emphasize the technical distinction that we introduced when discussing Marschak, between the case of *risk-love*, corresponding to that of a convex u function, and the case of *risk-aversion*, corresponding to that of a concave u . These two possibilities are equally compatible with the individual's preferences satisfying the VNM axioms, which, and which alone, constitute the VNM rationality criterion. The step accomplished by *Theory of Games* is to redefine "the problem of rational behavior" in such abstract terms that the concavity of u is no longer an analytic component of its solution, as it was in Marshall. A major consequence for the progress of economics, it now becomes possible to maintain that it is the science of rational conduct and nonetheless subject several forms of gambling and related activities to its theoretical investigations.

Perhaps mostly with a view of illustrating the new possibilities, Friedman and Savage propose, as a typical shape of utility function on monetary amounts, that u is concave on both small and large amounts of money, and convex in between. They suggest that this hypothesis "rationalizes" (1948, p. 293) – presumably, in the two senses of explaining and making rational sense of – the broad fact that the same individuals gamble and take out insurances, although for different amounts of money and different levels of income. On the face of it, this was an empirical hypothesis to be tested, and there was a detailed experimental study by Mosteller and Nogee (1951) – one of

the first ever made on EUT – which concerned itself with the proposed shape of u . The broad conclusion was that the suggested shape was not supported by the data.³⁶ Despite its hand-waving style, the hypothesis deserves to be remembered as an example of the new theoretical standpoint that emerged in the early 1950s.

We may conclude as follows on the formative stage of VNM theory. Prior to it, economics had little, if anything, to say about choices under risk in general. These choices could not be investigated properly because, by their very nature, they encompassed the excluded case of gambling-like behaviour as a relevant possibility. All of a sudden, thanks to von Neumann and Morgenstern, economists felt capable of exploring insurance behaviour, portfolio choices, job search, the firms' or the government's policies in a risky environment, and similarly important topics. Perhaps even more than the first readers of *Theory of Games* could have guessed, the 20 or 30 years following the book witnessed a continuous flow of deductions, usually from the simple VNM groundwork, but sometimes also from more refined theories like Savage's. Thus, with appropriate initial conditions and auxiliary assumptions, VNM theory was shown to entail definite statements on the agents' degree of insurance coverage, leading at long last to some elements of insurance theory.³⁷ A few – not all – of these entailments can be construed as empirical predictions, which testifies to the leap forward made in both logical and testable content. In retrospect, it seems difficult not to approve of the strategy pursued by the contemporaneous economists, like Marschak or Friedman and Savage, of protecting the fragile developments from *prima facie* counterexamples and foundational queries.

4 The Critical Stage of Expected Utility Theory

4.1 *The 1952 Turning Point*

The critical stage of EUT can be dated back to some events of the year 1952, after which the field was forever divided between its supporters and adversaries. First of all, a major conference gathering both North-American and European specialists of decision theory took place in May in Paris,³⁸ and intense controversies burst out within this distinguished attendance, which divided into roughly three groups. The American school, led by Marschak, Samuelson and Savage, was pushing the case for VNM theory all the more energetically since it could now rely on the sophisticated EU variant that the Savage had devised for subjective probabilities.³⁹ In this campaign, they received the idiosyncratic support of a founder of the subjective probability school, de Finetti. They were a powerful team, and in retrospect, one is struck by the boldness of the French economic engineers – most prominently, Allais – who pugnaciously objected to EUT throughout the conference. The majority of remaining participants were econometricians or mathematicians, who acted like witnesses – occasionally, like arbitrators – in the conflict between the bull-doing majority and the restless minority.

The year 1952 witnessed a second event of significance, which is not as straightforward to locate as the first. From then on, VNM theory would be discussed primarily in terms of a *privileged format*: a subject is asked to choose in succession between two pairs of lotteries with monetary outcomes, i.e., (x_1, y_1) and (x_2, y_2) ; the probability and outcome values are fixed in such a way that if the subject obeyed the VNM theory, he would choose x_2 if and only if he chose x_1 . The two-pair format is especially relevant to the testing and normative appraisal of the VNM independence condition (A3), whereas we have seen that a single-pair format is sufficient to discuss the continuity condition (A2).⁴⁰ Strange though it seems, the first sustained empirical attempts at testing VNM theory did not use the apparently obvious check that consists in presenting subjects with two (or more) relevant pairs. Rather, they implemented the following sequence: first, let subjects choose between lotteries $((p, x), (1-p, 0))$ with variable money outcomes x and probability values p ; second, estimate utility functions $u(x)$ from the choice data; third, test VNM theory by checking whether $Eu(x)$ accounts for these or (preferably) other data sufficient well. Why this “painstaking procedure” – the expression of a contemporary experimenter, Camerer (1995, p. 621) – lingered on to such an extent is a question we leave for the final assessment section.

The most famous (and indeed a very telling) example of a two-pair experiment is the so-called *Allais paradox*, which occupied centre stage in the experimental and theoretical developments of the 1970s and 1980s. The word “paradox” suggests connections with upsetting discoveries such as Russell’s paradox in set theory or the EPR paradox in physics, but there is no deep significance to be attached to it now; it is used simply because the subjects in Allais’s experiment usually choose against VNM theory, the established doctrine – the “doxa” – of the field. The history of this example is a genuine curiosum.⁴¹ It occurred for the first time during the 1952 conference, but is reported nowhere in the proceedings.⁴² According to a story which is now part of the decision theorists’ folklore, Allais asked Savage to choose from two pairs of lotteries when they were lunching privately, and his astutely chosen figures trapped him into a violation of VNM theory. Savage contributed to popularize the episode by revisiting it in his *Foundations of Statistics* (1954–1972); we will see how he made things worse for himself by trying to explain his blunder. The same figures appeared in Allais’s 1953 *mémoire* and the abridged journal form of this piece, also out in 1953, which is now the standard reference for the paradox.

Allais tried to debase VNM theory in two further ways, one of which is another paradoxical application of the two-pair format, and the other follows a different pattern, in which the damaging consequence is an inconsistency in measurement. These findings also belong to his work of the years 1952 and 1953. All in all, decision theorists were more impressed by the critical insights of this work than by the effort it also made to build a positive theory of risky choice. Following this line, we will discuss Allais’s paradox and the two lesser known paradoxes at some length. We then proceed to explain how they were turned in the 1970s and 1980s into the genuine experiments which they initially were not, and the rest of the section reviews some of the non-VNM hypotheses of the time, with a view to prepare the final assessment of VNM theory.

4.2 The Allais Paradox

In the journal article resulting from his *mémoire*, Allais (1953b, pp. 526–527) claims to have elicited answers to the following two questions:

Question 1: Which lottery would you choose of $x_1 =$ to receive 100 millions FF with probability 1,

and $y_1 =$ to receive 500 millions FF with probability 0.10, nothing with probability 0.01, and 100 millions FF with probability 0.89?

Question 2: Which lottery would you choose of $x_2 =$ to receive 100 millions FF with probability 0.11, and nothing with probability 0.89,

and $y_2 =$ to receive 500 millions FF with probability 0.10, and nothing with probability 0.90?

The numbers are so devised that a subject obeying VNM theory should choose either x_1 and x_2 , or y_1 and y_2 . For there is no numerical function $u(x)$ that would both satisfy the inequalities corresponding to the choice of either x_1 and y_2 , or of y_1 and x_2 :

$$u(100) > (\text{resp. } <) 10/100 u(500) + 89/100 u(100) + 1/100 u(0)$$

and

$$11/100 u(100) + 89/100 u(0) < (\text{resp. } >) 10/100 u(500) + 90/100 u(0)$$

Most very cautious people (“la plupart des gens très prudents”, *ibid.*), Allais contends, will choose x_1 and y_2 .

As Allais thought of the example, it affected (A3) rather any other part of VNM theory. To clarify the further step involved here, let us introduce the auxiliary lottery z :

$z =$ to receive 5 millions with probability 10/11, and nothing with probability 1/11,

and then reformulate the experiment as follows:

Question 1’: which lottery would you choose of $x'_1 =$ to receive 100 millions FF with probability 11/100 and 100 million FF with probability 89/100,

and $y'_1 =$ to receive lottery z with probability 11/100, and 100 millions FF with probability 89/100?

Question 2’: which lottery would you choose of $x'_2 =$ to receive 100 millions FF with probability 11/100 and nothing with probability 89/100,

and $y'_2 =$ to receive lottery z with probability 11/100 and nothing with probability 89/100?

The newly introduced lottery x'_1 and y'_1 have a common outcome occurring with a common probability – a *common consequence* in the later terminology –, and so do x'_2 (this lottery is unchanged) and y'_2 , the common probability being identical in the two pairs. Now, (A3) says that it is inessential to an individual’s preference how a common consequence is fixed; hence, he should strictly prefer x'_1 to y'_1 if and only if he strictly prefers x'_2 to y'_2 , and what he violates in VNM theory is precisely (A3). To draw this conclusion for the initial design requires one’s accepting *the axiom of reduction of compound lotteries*, which turns y'_1 and y'_2 into equivalent forms of y_1 and y_2 (the

rewriting of x_1 into x'_1 is unproblematic). Allais did not have to worry about the other axioms (A1) and (A2), which cannot be violated in this two-pair example.

Today's retrospectives often distort the initial intent of the paradox by describing it as a straightforward empirical test, which it was not and could not be from Allais's perspective.⁴³ When he first wrote about it, he was unable to document the prevalence of each answer; presumably, he had just experimented casually, as in the Savage episode. Besides, his article emphasizes, and it will then become a theme of his work, that *only rational peoples' answers matter*. Whatever evidence Allais might have gathered, the article claims it only for subjects who are, for one hand, rational by common consent (1953b, pp. 525, 528), and for another, well acquainted with the probability calculus (p. 524). It is because of the last assumption that Allais takes the compound lottery axiom for granted, apparently not noticing that this axiom can fail for other than cognitive reasons. He did undertake more thorough empirical research than his 1953 article suggests, but always limited to subjects exhibiting the above features, in practice a handful of high-brow students and colleagues.⁴⁴

Putting the paradox further in context, we find that Allais employs it *as a step in a normative argument*. Throughout the Paris conference, he strongly assumed that the EU hypothesis was false as an empirical generality, and even more strongly, claimed that his theoretical opponents shared this view, thus narrowing the disagreement with them to the issue of rationality.

Everybody recognizes the fact that man in reality does not behave according to the principle of Bernoulli [= the EU hypothesis]. There does exist a profound difference, however, in points of view as to how a rational man ought to behave. According to the American school, a rational man must conform to the principle of Bernoulli. In our view, this is a mistake (1953b, English summary, p. 504).

In order to advance the remaining issue of rationality, Allais – and this will also recur in his work – developed two strategies simultaneously. One amounts to arguing against EUT, and especially VNM independence, from the very meaning of the rationality concept. The other consists in *observing what rational men do*, and it is at this juncture that he introduced the paradox. Given the special argumentative purpose, the lack of empirical detail to fill it out is perhaps not surprising. The experiment, if this concept applies at all here, is *a thought experiment*, and it is arguably not even an empirical test, just the hint of it. The point is worth emphasizing, not only because it refines the received history of EUT, but because it creates an interesting tension for the philosophy-of-science reconstruction to come. We will have to check whether a Duhemian account can possibly register Allais's focus on the normative.

One should not make too much of the difference between the semantic argument about rationality and the thought experiment properly. Since the latter leaves undefined what characterizes “a man considered as rational”, one naturally turns to the former for clarification. Allais offers an “abstract definition of rationality” hinging on two main components: ordering and the dominance principle that was introduced in last section (roughly speaking, if, compared with l' , l increases the probabilistic weight of the better outcomes at the expense of the worse outcomes, l is preferred

to I').⁴⁵ Arguably, this minimalist definition fits in both with commonsense and the tradition of economics, which insists on not stretching rationality beyond mere consistency (1953b, p. 519). With rationality so conceived of, the VNM independence condition appears to be a gratuitous addition.

There is another line about *prudence*, which should be kept distinct from the argument about rationality, because it is not so much normative as psychological. Evidently, someone who selects x_1 from the first pair must be very prudent since he gives up x_2 , which has a much higher expected monetary value, just because of a 1/100 chance in this lottery of getting nothing. VNM theory would rationalize this choice by attributing an extremely concave function $u(x)$ to the subject; but then, $Eu(x)$ would predict the selection of y_1 in the second pair. That is, for this theory, the difference that a 1/100 chance of 1 million FF makes to the subject is the same whatever the other chances and outcomes may be. By maintaining that a cautious subject can select y_2 without falling into an inconsistency, Allais both rejects the standard theorizing in terms of expectation and concavity, and makes room for a novel conception in which *prudence would manifest itself variably depending on the particular pair of lotteries*. The same 1/100 chance of 1 million FF does not weigh the same when certainty is an option (first pair) and when both options are uncertain (second pair). In other, still Allaisian terms, there are *psychological complementarities between the chances*, which may be very strong when chances add up to certainty and quite weak otherwise.

The supporters of VNM theory were unshaken by these arguments. Unlike Allais, they did not have, nor wished to have, an explicit definition of rationality, and essentially reproduced Marschak's – after all classic – mode of reasoning in semantic circles from rationality conditions to examples, and vice-versa. Presumably, they thought that rationality was too elusive a concept to be clarified in the heavy style of a French engineer. It could only be discussed at the level of elaborate intuitions, and since these were far from being decisive, extraneous considerations, such as tractability and economy of description, had to be resorted to. Thus, it was seen as a decisive advantage of VNM theory that it summed up the individual's risk attitude in the shape of his $u(x)$ function, while preserving the convenient linearity of V in terms of probability values. This argument had widely circulated even before it served as defence against the Allais and related counterexamples.⁴⁶ In the *Foundations* (1954–1972, pp. 101–103), Savage tried to buttress the orthodoxy from still another side, claiming that only reflective choices counted normatively and that the paradoxical answers were no such choices.⁴⁷ Himself trapped by Allais's questions, he claimed to have restored preferences satisfying (A3) when he had noticed that he was violating it. He had made a mistake and corrected it; his thoughtful revision was an indication that he was at last answering rationally. But if a thoughtful revision points towards rationality, so does a thoughtful adherence to the paradoxical choice under the same informational circumstances, and Savage's alleged condition for rationality may well work like a boomerang.

4.3 *Experiments on the Common Consequence and Common Ratio Effects*

The Allais paradox would have remained a theoretical curiosum if it had not been for the efforts of a few decision theorists and psychologists to turn it into something like a controlled experiment. The first studies to be carried out, in the late 1960's and the early 1970's, involved asking variants of Questions 1 and 2 to a small group of subjects – typically, university students, but sometimes business people. The upshot was that the paradoxical choices occurred in 27% to 42% of the cases (these figures apply nearly as well to the violation of VNM theory, since the choice of x_2 and y_1 hardly occurred). One study allowed a subgroup of subjects to make their choices while discussing them, another involved the subjects' reconsidering their choices in view of the pros and cons stated by the experimenter, and a third had each subject assess the choices of others. In this way, the experimenters were paying some attention to Allais's concern about rationality. Above all, they were turning Savage's criterion of normative adherence against his favourite theory, for the paradoxical answer was the most frequent response in all types of sessions and even turned out to be slightly reinforced in one experiment involving reconsideration.⁴⁸

A sophisticated study was later realized by MacCrimmon and Larsson (1979) to double-check both the empirical prevalence of the paradox and its robustness to critical reflection. On the former score, they replaced some of the numerical figures in Questions 1 and 2 by *parameters*, which they let vary on a range of values that preserved the abstract structure. (i.e., Question 1 is the same as Question 2 but for a change of common consequence: see the analysis of last section.) This resulted in a rate of violation of about 30%, thus close to the initial findings. On the latter score, MacCrimmon and Larsson introduced the concept of a *rule* in order to capture the subjects' underlying motivations. Stated abstractly, but in natural language, rules would offer an educated subject a means of ratiocinating about his choices. Some were mediating the formal conditions of VNM theory, while others generalized the paradoxical answers directly. The idea was to compare the subjects' answers to choice problems with *their approval or disapproval of rules bearing positively or negatively on these choices*. Thus, MacCrimmon and Larsson asked their 19 subjects to give a score to each rule they presented, and eventually classified them into a consistent and an inconsistent group. Roughly speaking, with a consistent subject, the rules he obeys score better than the rules he violates, and the opposite with an inconsistent subject. The first group deserves more attention than the second, since its choices appear to have been more thoroughly reflected (and on Savage's criterion, to be more rational). In this privileged group, MacCrimmon and Larsson (1979, p. 368) noted a significant minority of paradoxical choosers.

This result may seem to be only mildly unfavourable to VNM theory; after all, a majority of consistent subjects were implicit EU maximizers. But if one interprets Allais's claim against the theory as being a barely *existential* one, as the context of his discussion makes plausible, this half-baked result is more than is needed. Besides, it should be recalled that Allais expected paradoxical answers only from

the very cautious among the rational people. In other words, the majority of MacCrimmon and Larsson's consistent group can possibly comprise of individuals who would depart from VNM theory in circumstances where a paradoxical choice would not require such a large amount of prudence as in the Allais-style examples.⁴⁹

Allais's *mémoire* and ensuing paper (1953b, p. 529) contains another thought experiment that is no less instructive than the celebrated paradox.

Question 3: Which lottery would you choose of $x_3 =$ to receive 100 millions FF with probability 1,

and $y_3 =$ to receive 500 millions FF with probability 0.98 and nothing with probability 0.02?

Question 4: Which lottery would you choose of $x_4 =$ to receive 100 millions FF with probability 0.01 and nothing with probability 0.99,

and $y_4 =$ to receive 500 millions FF with probability 0.0098 and nothing with probability 0.9902?

Arguably, some rational and prudent people choose x_3 and y_4 . This is another violation of VNM theory and – Allais contends again – of (A3) specifically. In view of the compound lottery axiom, the example is restated as:

Question 3: unchanged,

Question 4': Which lottery would you choose of $x'_4 = x_3$ with probability 0.01 and nothing with probability 0.99,

and $y'_4 = y_3$ with probability 0.01 and nothing with probability 0.99?

As it now appears, the violation of (A3) is obtained by offering in the second pair each alternative of the first with the same chance 0.01 and nothing otherwise. So Question 4' results from Question 3 by applying a *common ratio* – the later decision theorists' expression –, which is chosen so as to revert the choice, and thus plays the same disturbing rôle as the cleverly chosen common consequences in the Allais paradox.

Formally, common consequence and common ratio examples complement each other nicely, in that one group contradicts the statement that for given a and all r, r' ,

$$ap + (1-a)r \text{ R } aq + (1-a)r \text{ if and only if } a'p + (1-a')r \text{ R } a'q + (1-a')r.$$

and the other contradicts the statement that for given r and all a, a' ,

$$ap + (1-a)r \text{ R } aq + (1-a)r \text{ if and only if } a'p + (1-a')r \text{ R } a'q + (1-a')r.$$

Allais actually includes certainty as an option in the first pair, but this is really a particular case of the more general patterns exposed by these formulas.

MacCrimmon and Larsson (1979, pp. 354–359) also subjected the common ratio example to a parametric analysis, and varying the numerical values, found a rate of paradoxical choices reaching high values, although this required that lotteries promised large monetary outcomes. As before, they connected the brute responses with possible ratiocinations in terms of rules, and paid special attention to consistent subjects. Again two kinds of individuals were found in this group, implicit EU maximizers and paradoxical choosers, though not in a very large number for the last subgroup; however, the very fact that it is non-empty counts as evidence on the above interpretation of Allais's claims.⁵⁰

All in all, the common consequence and common ratio *effects*, as they are now classified, have been replicated a fair amount of times. The most telling results are perhaps those of Kahneman and Tversky (1979), who found violation rates in the region of 50% to 66% for both, when they asked questions with realistic monetary and non-monetary outcomes. The numerical figures in their two-pair experiments appear to be chosen even more efficiently than in earlier ones, supposing that the aim was to refute VNM theory *as an empirical generality*, but we have argued that MacCrimmon and Larsson responded to Allais's normative inspiration and thus did not conceive of a refutation exactly in this way.

4.4 *An Alleged Crucial Experiment: the Utility Evaluation Effect*

Another experimental scheme raised to popularity belatedly, despite having been suggested by Allais as early as the 1952 Paris conference. The uniqueness part of the VNM representation theorem severely constrains the functions $u(x)$ entering the EU representation obtained in the existence part. Once a zero point and an unit interval have been fixed for $u(x)$, this function becomes absolutely unique (whence the textbook comparison of the VNM index with a temperature scale). In particular, supposing that an individual satisfies the VNM axioms, *the same function* should result if two procedures based on the same zero point and unit interval are applied in succession to construct $u(x)$ from overt choices. A discrepancy between the resulting functions is evidence of the so-called *utility evaluation effect*. The test is crisp, and when proposing it, Allais went as far as to claim that it was “crucial” (p. 247 of the 1953 proceedings).

Allais himself investigated it experimentally by considering two procedures in turn. One – which is now said of the *certainty-equivalence* type – consists in revealing the sure money outcomes that the subject treats as being equivalent to given lotteries, and the other – now said of the so-called *probability-equivalence* type – goes in the opposite direction, thus revealing the lotteries that the subject treats as being equivalent to given sure money outcomes. Allais ended up with a most damaging conclusion for VNM theory, all of his 16 subjects exhibiting the utility evaluation effect (the report is in Allais, 1979, pp. 612–613, p. 634). Karmakar (1974) and McCord and de Neuville (1983) investigated the effect afresh with equally clear-cut results.

The fact that this effect is without exception, unlike the consequence and common ratio effects, makes it crucial in the loose sense, which we have found to be used by Popper, of an empirical result that refutes some theoretical compound indisputably. But it does not make it crucial in Duhem's sense of discriminating either between two theories, or between two component parts of the same theory. Quite to the contrary, the sharp conclusion against VNM theory *as a whole* is paid a large price in terms of confusion of possible sources of falsity *within* that theory. Logically, the candidates for violation are not only the VNM independence axiom (A3) and the reduction of compound lottery axiom, as in the previous effects, but also the continuity conditions

– related to (A2) – that underlie the way of constructing $u(x)$. To illustrate the point, take a procedure of the above described certainty-equivalence type: it would come to a halt if for some lottery presented to the subject, his answers pointed towards a sure equivalent higher than the best outcome provided by this lottery. This would be a violation of continuity formally identical to that of Marschak's rock climber in Section 3. One may rejoin that the utility evaluation effect *itself* would vanish for such a subject, but this would be tantamount to saying that the effect is not universal after all. Also, the construction of $u(x)$ hinges on inferring an indifference statement when the strict preference is reversed – e.g., if the subject claims to prefer 100 euros to a lottery l and l to 99 euros, he is supposed to be indifferent between l and some amount between 99 and 100 euros, and this aspect of the design appears to involve part of the ordering axiom (A1) in the interpretation of the experiment.

Neglectful of these logical possibilities, Allais and most writers after him have treated the utility evaluation effect as another violation of (A3), on a par with the common consequence and common ratio effects. Once again, Allais resorted to the argument that complementarity of chances becomes irresistible to the prudent when these chances add up to certainty. To illustrate, consider two procedures of the certainty-equivalence type that differ by the first lottery they present to the subject. One may intuitively expect that they will entail distinct $u(x)$ *if the lottery is close to certainty for the first, and genuinely uncertain for the second*.⁵¹ Now, since procedures of the probability-equivalence type depend on offering sure outcomes, not lotteries, nothing in the argument lets one expect such a strong dependence on the first stage, so it eventually suggests that the two types will be at variance. Here as with the other effects, the experimentalists' work has filled Allais's theoretical guesswork with substantial evidence.⁵²

4.5 *The Experiments in the Light of Alternative Theorizing*

The late 1970's and early 1980's witnessed not only the counterevidence to VNM theory that we have gone through, as well as some more that we have not, but also theoretical construals that aimed at accommodating these data, and for the more ambitious ones, at founding the theory of risky choice on a novel basis. By the end of our period, one of the contributors, Fishburn (1988), could mention about a dozen of purported alternatives to VNM theory beside his own. Not all of them are relevant to the present study, which is only concerned with investigating the sense in which refutations occurred, and Duhemian underdetermination was overcome, with respect to this theory. Thus, with due apologies to some of the inventors, we drastically restricted the sample of this paper to four proposals, i.e., Allais's theoretical work up to 1979, Kahneman and Tversky's (1979) "prospect theory", Machina's (1982) "generalized expected utility theory" (GEUT), and "rank-dependent expected utility theory" (RDEUT), which can be traced back to several contributions of the 1980's, the earliest one being Quiggin's (1982).⁵³ The first two contributions are unavoidable, because, as their dates and authors suggest, they might have

helped to *shape* the counterevidence against VNM theory. By contrast, the third and fourth contributions are not coincidental with, but reflective upon, the production of counterevidence. They have been included here because each is representative of a major trend of revision – GEUT offering something like a conservative refinement of VNM theory, while RDEUT breaks away from it in significant respects. Some of the received answers to Duhem’s underdetermination problem involve attributing a rôle to superseding theories, so there is a philosophical reason for extending our inquiry somewhat in this direction.

Further clarifications are in order before we proceed. First, with typical looseness of terminology, the purported alternatives to VNM theory have been referred to uniformly as “theories” although quite a few of them consisted of a rather thin package – a single mathematical formula, plus some informal motivation and fairly scattered evidence. A philosophical (and indeed any careful) treatment should attend to the distinction *between a specific hypothesis and a full-fledged theory*, so that we will in this case refrain from using the vocabulary of the field, despite the inconvenience. Second, there are many ways in which a positive alternative – however sketchy – can relate to a refuting test. *Ex ante*, it can provide it with an abstract scheme, as Allais’s argument about complementarity of chances illustrated. *Ex post*, it can offer a heuristic account of the results, while perhaps connecting them with those of previous tests – we have also seen this at work with Allais. The account can possibly reach the level of a proper explanation, but we have met no evidence of this achievement at this stage.

With these warnings in mind, let us briefly examine Allais’s “positive theory of choices involving risk”. It actually consists of two distinct generalizations of VNM theory. The first by order of time precedence – it dates back to 1952 – replaces the EU formula with another weighted sum of utility values, in which the weights are *functions of all parameters under consideration*. Formally, if $I = ((p_1, x_1), \dots, (p_m, x_m))$,

$$V(I) = a_1 u(x_1) + \dots + a_m u(x_m), \text{ with } a_i = a_i(p_1, x_1, \dots, p_m, x_m), i = 1, \dots, m.^{54}$$

This new formula is not quite vacuous because of the separation between utilities and weights, but it is badly in need of specification. Allais did not flesh it out until 1988, when he rediscovered RDEUT, to be reviewed below. As early as 1953, he offered a second line of analysis, which prevails in his overall work. It proceeds from the claim that *the dispersion of utility values matters no less than their average value*, which the EUT formula only takes into account. Accordingly, the suitable generalization is either

$$V(I) = g(F(u(x_1), \dots, u(x_n))),$$

where $F(u(x_1), \dots, u(x_n))$ is the probability distribution of the possible utility values, or

$$V(I) = h(Eu(x), Var u(x), \dots),$$

where the second argument of h is the variance of the utility values, and the unspecified arguments represent as many higher-order statistical moments as one may wish to add.

The last two formulas had in fact spontaneously emerged during the Paris conference, in connection with both the VNM and Savage theories. They were obvious to the other mathematically-minded participants, and if Allais parted company with them, this only because he interpreted them differently. The formulas were indeed equivocal, since they denied EUT the status of an exact theory, but allowed for the possibility that it would be a decent first-order approximation. This last stance was taken by Frisch when he summarized the conference debates:

We must try to proceed further... by looking at... *approximations*... Take any choice structure, say the structure of Paul. In reality, this structure will most likely not satisfy the neo-Bernoullian axioms exactly, but it might do so approximately (pp. 253–254 of the proceedings).

Allais emphatically rejected such a soothing view of the performances of EUT. One of his arguments appears to be a priori in the strict philosophical sense. The dispersion of utility values is allegedly “the specific element of the psychology of risk” – a necessary part of any risk-attitude, to paraphrase. It follows that a first-order approximation of the V function cannot be correct in general; only a second-order one can conceivably be.⁵⁵ The Allais paradox comes to the rescue of the a priori reasoning. Here is indeed a thought experiment to suggest that the first-order approximation fails miserably, and by contrast, that the generalizing formulas work.

Given that Allais puts so much weight on the psychological complementarity of chances, one would expect him to represent it in the present framework, but at this juncture, he is distressingly vague. Clearly, this argument requires that $V(l)$ be non-linear, which the equations permit, but the point is to have the equations *entail* it, and for that, the shapes of g and h must be specified. When this is done, one will also be able to answer two important questions left pending, i.e., for what kind of choices the first-order approximation fails – since from Allais’s own admission, it does not always do – and whether the second-order approximation is sufficient in general, or in turns calls for refinements. Only in 1979 (see p. 482 and Appendix B2) does Allais become more precise, but not to the point of answering these queries fully. He puts forward the still very general formula for g :

$$V(l) = Eu(x) + H(f(u(x) - Eu(x))),$$

where f is the probability (or density) function of utility values, and H an arbitrary function.⁵⁶ Going one step further, one of Allais’s followers, Hagen (1979), stated h thus:

$$V(l) = Eu(x) + \alpha \text{Var } u(x) + \beta [E(u(x) - Eu(x))^3],$$

with $\alpha < 0$ and $\beta > 0$. As the sign of the coefficient indicates, the individual is supposed to be averse to the dispersion, but this aversion is asymmetrical, since it depends on how the good outcomes (those above $Eu(x)$) compare with the bad ones (those below $Eu(x)$). At long last, Hagen’s formula answers the queries.⁵⁷

With this formula, the Allaisian school might have reached the appropriate degree of theoretical specification, but it would still be open to another criticism: it does not have theoretical foundations in the same sense as its EU competitors.

Allais often claimed to have provided “axioms”, but he just meant by that unifying assumptions from which diverse consequences follow. His work falls short of the standard of axiomatics set up by the followers of von Neumann and Morgenstern, and above all, Savage, which may explain why the latter objected to him that he had no real alternative to offer. Allais’s “axioms” involve the utility function $u(x)$ as a datum, whereas Section 3 explained that these writers made a point of deriving it. Remarkably, this proves to be not only a theoretical weakness, but *a problem for the experimenter*. If one estimates Hagen’s equation directly, one is left with too many free variables for the same set of choice data, and the experimentation risks becoming a curve-fitting exercise (see Camerer, 1995, p. 627). The advantage of axiomatizations à la VNM-Savage transpires at this point, since they analyze the utility representation into more elementary, qualitative properties that are amenable to a test without the problematic phase of estimating $u(x)$.

In sum, Allais’s “positive theory” is far from constituting the alternative to EUT that he and his disciples claim it to be, and for a long time, it has not even been a theory at all, but rather a heuristic to find one. Even in its final form, it is not clear whether it *explains* the empirical evidence or is but a redescription of the latter. If it matters at all, this is because it contributed to suggest the famous paradoxes, over and above the informal arguments that might have been sufficient for the task.

The second work in our sample, Kahneman and Tversky’s (1979) “prospect theory” briefly emerged as a plausible alternative to VNM theory until some of its weaknesses became apparent. It develops in two parts corresponding to two successive stages of an idealized choice process, which are called *editing* and *evaluation* respectively. An objectively posed choice problem is restated subjectively at the first stage and resolved in this adapted form at the second stage. Kahneman and Tversky’s paper is justifiably a classic, but a good deal of its novelty lies in its emphasis on editing, and this part virtually eschews comparison with standard decision-theoretic reasoning. Those concerned with improving on VNM theory have generally focused on the evaluation part, which follows more or less classical lines. Here, the two psychologists propose replacing the probabilities p_i by weights of the simplest possible functional form:

$$V(l) = a(p_1)u(x_1) + \dots + a(p_m)u(x_m).$$

The outcomes are monetary, and for simplicity, the authors focus on lotteries having at most two non-zero such outcomes. Some experimentalists of the 1950s and 1960s, like Edwards, had proposed the same formula in response to the first apparent violations of VNM theory, generally interpreting $a(p)$ as a *subjective distortion* of the given probability vector p .⁵⁸

Kahneman and Tversky rejuvenate this line by making new hypotheses of about the u and a functions. As regards the former, they conjecture that $u(x)$ is defined on *changes* in wealth rather than total wealth, and that this function exhibits risk-aversion (concavity) on the range of gains and risk-love (convexity) on the range of losses. Since 1979, empirical support has built up in favour of this conjecture, which has in effect displaced Friedman and Savage’s (1948) failed attempt at determining the curvature of u . As regards the latter function, Kahneman and Tversky

begin by assuming that a is increasing between the extreme values $a(0) = 0$ and $a(1) = 1$, and then propose substantial restrictions, *each of which is dictated by a well-established effect*, which makes their technical discussion exemplary. One of these restrictions, i.e., that low probabilities are overvalued and high probabilities undervalued, directly connects with the common consequence effect. They also deal with the common ratio effect, and if they do not envisage the utility evaluation effect, this is because it was little known at the time. Having checked for the logical compatibility of the shape restrictions, they eventually suggest to take a to be convex except in the vicinity of 0 and 1. This probability-related analysis has not gained the same acquiescence as the utility analysis of the paper, in part because RDEUT, which defines a differently, has superseded “prospect theory” in most specialists’ opinion, in part because those who have adopted the same definition have sometimes favoured other shapes.⁶⁰

The previous formula for $V(I)$ permits violating the dominance principle, and this is the main reason why the “alternative account” that Kahneman and Tversky’s (1979, p. 263) announced did not really take off. Once this possibility was demonstrated, most decision theorists lost interest, because they strongly believed that a proper alternative to VNM theory should retain the principle (in this respect, Allais is representative of the community as a whole). Kahneman and Tversky (1979, p. 283) had actually guessed that a non-linear a conflicted with dominance preference, but they had responded by making the editing stage responsible for the detection and discarding of dominated lotteries (*ibid.*, p. 275). This is an extraordinary move if one thinks of it, since most lotteries are dominated by others in a standard VNM setting, so that the evaluation stage would be left with near to nothing to operate on.

Independently of the conundrum created by dominance, those accustomed to EUT axiomatizations could complain against “prospect theory” on the formal grounds that it lacked foundations. The evaluative part revolved around a single formula with utility as a datum, a formula which enjoyed some inductive, but no deductive support, and was not axiomatized. In fact, Kahneman and Tversky had proposed a specific hypothesis rather than a full-fledged theory, which in part reflected the different working habits of decision psychologists and mathematical decision theorists.

In retrospect, leaving aside the new definition of u , “prospect theory” matters because it sharpened and helped to reorganize the empirical counterevidence of the time to VNM theory (though not the utility evaluation effect). Also, it signalled a major shift in the understanding of experiments, which henceforth became directed at refuting or confirming empirical regularities instead of entering into a normative assessment of rules of conduct, as was the case in the work closely following Allais.

The third potential alternative to be considered here, Machina’s (1982) “generalized expected utility theory” (GEUT) had a twofold aim, for one to weaken VNM independenceso as empirically to account for several known effects at once, and for another, to preserve as much as possible of the apparatus developed by VNM theorists to deal with risk-attitudes. The strategy was to strike a balance between

conservatism and revision, as it were. In terms of the axiomatization of Section 3, it consists in keeping (A1) and (A2), while replacing (A3) by two conditions, (H1) and (H2), that are directly stated in terms of the V function.

(H1) The function V representing \leq over L is differentiable in the probabilities.

Barring technicalities, it is clear what this condition means to achieve: while linearity in the probabilities is a global and exact property, replacing it with differentiability will preserve it *locally and as an approximation*. Frisch and others had pointed out this way out, and Machina is now following it to the end. A precise formulation of (H1) requires mathematical care because of the non-standard domain L ,⁵⁹ and Machina simplifies his task by assuming that the set of outcomes X comprises of all possible numbers between 0 and some positive M . This is sufficient if the aim is to handle experiments with money outcomes as well as the more standard economic applications. It follows from (H1) that for any chosen reference lottery l , there exists a function $ul(x)$ entering a linear formula that approximates V around l . Thus, something of the VNM representation theorem can be salvaged, as one could hope. Further – a most useful contribution – the VNM analysis of *preference for dominance* and *risk attitudes* turns out to be essentially preserved. The first step is to check that these objects can be investigated locally as they were globally, i.e., in terms of the monotonicity, concavity and convexity of the ul . Then comes a second, more ambitious step, which is to reconstruct global properties in terms of local ones, and then show that some of the existing theorems obtained under VNM theory still hold for GEUT.⁶¹

Besides its applications to economics proper, for instance to insurance theory, the preceding analysis permits handling the experimental effects of this section, as well as some more. By suitably adjusting the curvatures of the ul , one can make V compatible with choices that are sometimes cautious, and sometimes not, as in the common consequence or common ratio effects. But clearly, it would be too easy a game to fix the local utility functions just as the subjects' answers require, and this is where the supplementary condition (H2) comes in. It takes into account a *systematic* property of the observed violations that we have not yet brought out. In Allais's Questions 1 and 2, x_1 dominates x_2 , and y_1 dominates y_2 , in the sense relevant to the dominance principle – that is, x_1 (y_1) gives more weight than x_2 (resp. y_2) to the more valuable outcomes. Other experiments performed on the common consequence effect obey this pattern, which can be formalized thus: in the sense made precise by the concavity of his ul functions, the individual becomes *less cautious* when he moves from a choice between lotteries x_1 and y_1 to a choice between x_2 and y_2 such that x_1 dominates y_1 and x_2 dominates y_2 . As Machina (1983) explains, the same pattern of variation occurs with the common ratio and the utility evaluation effect. This finding motivates the next condition.

(H2) If a lottery l dominates another lottery l' , ul will exhibit no smaller risk-aversion than ul' .

In the particular case of three monetary outcomes, (H1) and (H2) admit of an elegant two-dimensional geometric representation.⁶³ The individual's indifference loci are not parallel line segments anymore, as VNM theory requires, but continuous and smooth curves (because of (H1)), which "fan out", i.e. become steeper in

the direction of increasing dominance (in connection with (H2)). Given that (H1) makes such a wide generalization step, (H2) is essential to the determinateness of GEUT. Without it, one can doubt that this theory would explain anything at all. But even with it, does it really explain the paradoxes? Machina claims that “all follow from a single assumption [= (H2)]... which leads to further refutable restrictions on behavior” (1983, p. 282). This is correct only if one does not take “follow” in a straight logical sense. For (H2) does not tell *by how much* the curves become steeper, and lacking quantitative precision, it cannot logically entail, for instance, that an individual satisfying it will answer Allais’s questions in the paradoxical way. In fact, as the formal statements indicate, (H1) and (H2) together are compatible with the individual’s choosing in the VNM way. The sense in which GEUT *explains* the paradoxes must then be qualified thus: for any given pair of choices that violates VNM theory, if one assumes (H1), one can find a quantitative specification of (H2) that entails the violation. This sense of “explanation” involves an element of curve-fitting.⁶²

GEUT attracted, and still attracts, considerable interest from decision theorists, but few would describe it as a proper alternative to VNM theory, and the reasons for this appear to be twofold. For one, despite its axiomatic flavour, it does not consist of an axiomatic system in the received sense because it uses $V(I)$ as a primitive term. There are serious, perhaps insuperable, difficulties involved in the rendering of (H1) at the qualitative level of a preference relation. For another, and more importantly, (H2) is only weakly predictive and weakly explanatory, as was just explained. This is not to say that informative tests cannot be performed, because (H2) excludes some VNM configurations (those with “fanning in” or no fanning at all), and can anyhow be rejected by the data despite being weak. The experimental work along these lines has suggested a mixed record.⁶⁴ In retrospect, its significance may primarily rest with the generalizing step contained in (H1), which implements the idea that VNM theory can be stated as an approximation.

4.6 Rank-Dependent Expected Utility Theory

The last item in our narrow list, RDEUT is by now the best regarded, and it can indeed be argued that it deserves the highest consideration. Historically, it is a curious case of intellectual convergence, because its basic mathematical formula was worked out independently by Quiggin (1982), Yaari (1987), and Allais (1988), and these writers came across it starting from somewhat different motivations. Formally, take a lottery

$$l = ((p_1, x_1), \dots, (p_m, x_m)),$$

with the outcomes x_p, x_m being ranked in that preference order from the lowest to the highest. Then, the RDEUT evaluation is

$$V(l) = a_1 u(x_1) + \dots + a_m u(x_m),$$

where:

$$a_1 = f(1) - f(p_2 + \dots + p_m), \dots, a_i = f(p_1 + \dots + p_m) - f(p_{i+1} + \dots + p_m), \dots,$$

$$a_m = f(p_m),$$

and f is an increasing function from the 0–1 interval to itself that satisfies $f(0) = 0$ and $f(1) = 1$. Equivalently:

$$V(l) = b_1 u(x_1) + \dots + b_i (u(x_i) - u(x_{i-1})) + \dots + b_m (u(x_m) - u(x_{m-1})),$$

with $b_1 = f(1), \dots, b_i = f(p_1 + \dots + p_m), \dots, b_m = f(p_m)$

This formula generalizes that of VNM theory, which is recovered by putting $f(p) = p$. As the first equation makes clear, it consists of a weighted sum of utilities in which the weights a_i are functions of probability *sums*, not of individual probability values as in “prospect theory”. The sums compute the probabilities that the realized outcome will *reach or exceed* the successive preference ranks: for $p_1 + p_2 + \dots + p_m = 1$ is the probability of getting at least the worst outcome x_1 ; $p_2 + \dots + p_m$ is the probability of getting at least the second worst outcome x_2 ; and so on until one reaches the probability p_m of getting exactly the best outcome x_m . So what f transforms is the *decumulative* distribution (i.e., one minus the cumulative distribution, as usually defined) that is associated with the probability vector p .

The second equation suggests an informal rendering in terms of a sequential evaluation process. Taking each x_i in turn, the individual applies to its value $u(x_i)$ an abatement ratio b_i that depends on how uncertain he is of receiving x_i or a better outcome. He knows that he will enjoy at least the utility of x_1 with certainty, hence $b_1 = 1$, that he will enjoy at least the utility increment of x_2 with some uncertainty (whence $b_2 < 1$), and so on until he considers the utility increment of x_m , which he is maximally uncertain to enjoy (whence $b_m < \dots < b_2 < 1$).

RDEUT has a synthetic value that its inventors perhaps did not realize fully, and in any case did not stress, because they had different points to make in using the same mathematics. First of all, it solves the problem that had plagued “prospect theory”. For the above definition of $V(l)$ satisfies the dominance principle by the increasing property of f , and it is demonstrably unique in satisfying the principle within the class of functions having a similar weighted sum structure. This uniqueness property was Quiggin’s main justification for adopting RDEUT, and it is also a reason for Allais, who had toyed with weighted sums and also argued for the rationality of the dominance principle.

Second, the probability transformation function f comes with an interesting interpretation in terms of *risk-attitudes* that was not so obviously available with other functions in the same class. Quite naturally, the early experimenters tended to think of their $a(p)$ as of “subjective probabilities”, in the – non-Savagean – sense of a cognitive distortion of the given lottery probabilities. Thus, overweighting of small probabilities meant that the subject regarded the corresponding events as more probable than they were; and symmetrically for underweighting of large probabilities. With RDEUT, the dominant interpretation – fixed by Yaari and Allais – is that f reflects the individual’s propensity to take chances, not his misperception of their

magnitude. This semantic line is comforted by mathematical results that connects with the VNM analysis of risk attitudes in terms of the u function. In a sense that these results make precise, f and u are dual to each other.⁶⁵ Very roughly speaking, the risk-aversion property that VNM theorists identify by saying “the u function is concave” is now captured by saying “the f function is convex”. Quite a few theorists today have replaced one statement by the other, thus dramatically breaking up with the received analysis.⁶⁶

Last but not least, the RDEUT evaluation satisfies a weak form of VNM independence, the so-called comonotonic independence, which is easily interpretable and can be tested exactly to the extent than its ancestor condition can (i.e., with the same possibilities for confusion of possible sources of falsity). Two lotteries l, l' in L are said to be *comonotonic* if $l = ((p_1, x_1), \dots, (p_m, x_m))$ and the outcomes $x_1, \dots, x_k, \dots, x_m$ and $y_1, \dots, y_k, \dots, y_m$ are ranked by increasing preference order.

(A3⁻) (Comonotonic Independence) For all l, l' in L , and all numbers a between 0 and 1 (0 excluded), if l and l' are comonotonic and for some $k, x_k = y_k$, then $l R l'$ if and only if $l_* R l'_*$,
for all l^* and l'^* , which are identical to l and l' , respectively, except that $x_k^* = y_k^*$ with x_k^* possibly different from x_k .

Compared with (A3), this condition limits the replacement of a *common consequence* to those cases in which the initial lotteries are comonotonic and the replacement does not upset the order of outcomes in the lotteries. It is easy to check that the Allais paradox choice does not violate (A3⁻), which is well suited to handle the common consequence effect generally. The new axiom can be defended on the ground that l and l_* on the one hand, and l' and l'^* on the other, exhibit the same qualitative pattern of risk, so that there seems to be no reason why the preference between l and l' should differ from the preference between l^* and l'^* . This is the gist of a *rationality* argument for obeying (A3⁻) and not obeying the stronger (A3).⁶⁷

Comonotonic independence can be included in a full-fledged axiomatization of RDEUT, so as to reach the formal standard of VNM theory and make comparisons with it more definite. In the more advanced framework in which probabilities are not given, there exists a related condition, which is the comonotonic counterpart of a classic axiom in Savage's system, and this condition can similarly be embodied in a full-fledged axiomatization. Actually, the move on this front was initiated by Schmeidler (1986) independently of the move against VNM theory.⁶⁸ We have excluded Savage from our investigation, and must then gloss over an important development here, but it needs stressing because it is part of the collective discovery of rank-dependent evaluations, and conceptually, further proof of their theoretical significance.

From the viewpoint of empirical performance, RDEUT is open to the same problems as GEUT. The curve-fitting element is obvious from the solutions given by Quiggin and others to common consequence or common ratio problems. In general, given an observed choice in a two-pair experiment, concavity or convexity restrictions on both u and f are not sufficient to entail either this choice or its absence, and one should take the step of fixing some numerical values for both functions. The

comparative empirical record is mildly favourable to REUDT,⁶⁹ but it seems clear that this has played only a minute rôle in tilting the decision theorists' scales towards it – the main reasons being the convergence of arguments to support the weighted sum formula and the availability of an axiom system which articulates them coherently. The success of RDEUT is relevant to Duhem's underdetermination problem only on the view that the solution to this problem requires that a superseding theory be available, but more generally, it tells us much about the way decision theory evolves and perhaps makes progress.

5 A Quasi-Duhemian Account of the Historical Sequence

5.1 Counterexamples, Anomalies, and Effects

Right from its beginning, VNM theory was surrounded with *counterexamples*, a situation which its exponents were generally willing to acknowledge. The troubling cases were suggested by casual observation of human life and required ingenuity only to connect them with particular axiomatic conditions (such as continuity in adventurous behaviour and reduction of compound lottery in multi-stage gambling). Those who accepted and even provided the counterexamples conceived of them not as *refutations* of the theory, but as *directions to specify its domain of application* more precisely than the founders had done.⁷⁰ A common idea was to specialize the theory in economic applications, but this turned out to be a red herring. Attempts were made to revive the economists' exclusion of the pleasure of gambling, and in the final one, this led to a division of risk attitudes, some but not all of them falling within the competence of the theory. What is noteworthy here is that VNM theorists could not assign a domain without resorting to the rationality concept, for which they had no independent definition, and which they explored *by using the concepts of the theory itself*, a circular move that was perhaps unavoidable.

Anomalies are another particular case of refutations, bordering refutations, but not identical with them. They do not impact on any statements, but only generalities, and they enjoy the dialectical privilege of exceptions, which both infirm the rule and confirm it, provided they are neither too deep nor too numerous. Thus, anomalies can be seen as refutations in a virtual state that is realized only if they are reconceptualized or widely replicated or both. Since Kuhn and Lakatos, it has become standard to claim that scientific theories are accompanied with anomalies. However, the first counterexamples to VNM theory were *not even treated as such*, the first case being arguably the Allais paradox, which does not belong to the initial stock. Lakatos (1970, p. 120, n.2) famously claims that "in actual history new theories are born refuted; they inherit many anomalies of the old theory". This does not appear to apply here. One of the reasons is that VNM theory developed in something like a vacuum, because the previous economists had eschewed any serious investigation of risk attitudes.

The fact that there was so little done on this major topic before *Theory of Games and Economic Behavior* is the strongest argument in favour of the one-sided strategies of the early years. Had decision theorists yielded to the fascination of counterexamples, microeconomic textbooks would look very different today. Insurance economics, a good deal of theoretical finance, as well as many of the received game-theoretic applications to industrial organization, employment contracts, investment policies, or more recently auctions, took off thanks to the EU formula and the trick of representing the agents' risk attitudes in terms of the utility function. In order to protect these usable components from criticism, the early theorists defended VNM theory *as a whole*, despite the clear suggestion of the counterexamples that some axioms were more dubious than others. They were holists, but technically not Duhemians, since an analysis of domain restrictions is not to be found in *TP*. They were not Kuhnians or Lakatosians either, because, as just said, they had no use for the concept of an anomaly, to which domain restrictions are typically referred in the growth-of-knowledge literature.

The Allais paradox acquired the status of an empirical refutation without going through the intermediary stage of an anomaly, but this happened for reasons that parallel those which turn an anomaly into an empirical refutation. For one, a deeper analysis uncovered its significance, and for another, it was transformed into an empirical generality of some kind. The paradox became an *effect* (the common consequence effect), a metatheoretical concept of decision theory that may now be analyzed. Put abstractly, it is an empirical phenomenon that is endowed with some regularity, though it admits of exceptions, and is accompanied with some explanation that is both causally relevant and psychologically intelligible. There is often a semantic suggestion that the phenomenon can be reproduced with some success by the experimenter, which would imply that the causal explanation is adequate at least for manipulative purposes. An effect belongs to the outside world of observation, but typically receives its name from the factor that supposedly triggers it out. Mathematical decision theorists like Machina (1983) tend to identify effects with formal statements that enter logical relations with each other and with axiomatic conditions. Kahneman and Tversky (1979), and decision psychologists in general, are much less formal. They do not worry so much about entailment as about compatibility relations that support causal explanations. Also, psychologists often stop their explanations of choices at the level of an effect when it is sufficiently unexceptionable, while mathematical theorists regard them as being at best low-level generalities in need of a unifying deductive explanation.

Here is how this concept relates to our main theme: once a counterexample such as the Allais paradox is registered as an effect, one major item in the Duhemian compound – i.e., *the observational record* – disappears from consideration. In the particular instance, the registration process obeyed the pattern that Duhem let one expect for a solution to the underdetermination problem in general – there was a relatively lengthy exchange of pros and cons with an overall conclusion that the cons weighed more heavily. Indeed, the anti-EUT continuously brought out challenging facts, somewhat like the undulatory theorists, while the pro-EUT mostly limited themselves to reinterpret the evidence, somewhat like the corpuscular theo-

rists. At the risk of exaggeration, one may liken any of the three effects discussed in this paper to the lower-level generality that was at issue in Foucault's case, i.e., that under certain specifiable and reproducible conditions, light moves more quickly in the air than in water.

To claim that an effect, in the above sense, is reasonably well established presupposes that some check has been made of the possibly disturbing rôle played by the *concrete conditions* in which the experiments took place. This implies the more basic requirement that these conditions be identified with some objective precision. What were the physical devices employed to put the choices to the subjects, and especially, to convey the probability numbers to them? When the experimenter varied the pairs of lotteries, how many did he put in a row, and did he present each pair only once or several times, and in the latter case, after what time lag? Were the chosen lotteries drawn, and if so, were the resulting money outcomes effectively paid for all, some, or none of the drawings? Were the subjects students, businessmen, or ordinary people? Had they been exposed to the concepts of the theory before? When the experiment involved some discussion of the VNM conditions, at what level of abstraction was it conducted? In the period under consideration, the experimental reports were terse, and from what can be inferred, not all of the problematic conditions were identified, let alone controlled for. For instance, experimenters may have worried about the possibility that students reacted differently from ordinary subjects, but they nonetheless continued to use the same abstract questionnaires for the various social groups. They hardly addressed the tangled issue of real payments and what differences it made to introduce them, both in terms of stronger incentives and wealth effects changing the subject's greediness. These loose ends are easy to point out now that experimental economics circulates more or less standardized lists of factors to check.⁷¹ Still, the 1970s and 1980s witnessed a number of replications, with conditions changing *at least from one experiment to another*, and since the aim was to establish effects of a broadly qualitative kind, one may conclude that this unsystematic variety was sufficient. The utility evaluation effect can be said to be well supported, and the common consequence and common ratio effects, to be recognizable from the data, the former probably with a better record than the latter.

5.2 Confusion of Sources of Falsity Once the Empirical Record is Established

In Duhemian terms, once the step is taken to regard the observational record as unproblematic, the remaining sources of falsity are the primary hypotheses and *some* of the auxiliary hypotheses (a good deal of them having been handled at the previous stage). In VNM theory, the primary hypotheses are the preference axioms, which were tested by investigating choices in an artificial context of questions and answers. It would have been possible to register their spontaneous risk-taking choices, as Friedman and Savage (1948) coarsely did when they observed that the same people gamble and take out insurance. There is a threatening possibility that

the effects, although well evidenced among the subjects of an experiment, do not carry through to the actual, uninfluenced behaviour of nondescript individuals, and are thus irrelevant to the theory since it is *that* kind of behaviour which it means to capture in the end, whatever the precise frontiers of the application domain. Logically, the experiments that establish an effect are conclusive against VNM theory only under the supplementary assumption called *parallelism* or *external validity* in today's discussions of experimental economics. There is a clear contrast with the *internal validity* assumptions that were exemplified in the last subsection. To some extent, parallelism involves a metaphysical claim of the "uniformity of nature" or "limited variety" kind, which is so general as to underlie any experimental work, but it can also be viewed as making an empirical – and even possibly testable – claim on how humans behave under different environments. If the latter line prevails over the former, parallelism should be included in the list of auxiliary hypotheses which impinge on the conclusion that the effects refute VNM theory.⁷²

Importantly, work based on this theory has persisted up to now in economics, both at the applied and theoretical level, and this can be explained along various lines. One is the economists' well evidenced stubbornness when they meet facts that challenge their theoretical outlook.⁷³ But a perhaps more attractive explanation is that they are reluctant to accept parallelism in its empirical interpretation. For they are not impervious to *any* choice evidence, but only to that which does not take certain standardized forms. In particular, choices should be made in the context of a genuine economic activity, most typically one involving significant money flows, and they should be observed from the outside rather than reported in verbal statements. Some, though not all, economists would add that choices should be observed across a sufficiently large number of different individuals for the most idiosyncratic features to cancel out.⁷⁴ There is still another interpretation, which is exclusively applicable to the topic of this paper, i.e., that the reluctant economists take the view that VNM theory is inaccurate rather than plainly false, and should be preserved as an approximation for lack of a better alternative. This view was floated as early as the 1952 Paris conference, and we have seen that it connects with Machina's "generalized expected utility theory" (GEUT). Even if Machina has shown that some of the standard economic theorems still hold in his framework, other applications require the linearity of the EU formula exactly, not just approximately. Being more uniform, the previous suggestion that economists reject parallelism is perhaps the best suited to redeeming their work.

Both experimental and non-experimental applications raise the little noticed problem that the subject's *choices* are imperfectly related to his *preferences*, which are the true object of the axiomatic conditions. Identical in this respect, decision theorists and economists work with such an impoverished semantics that they ignore this possible discrepancy, but the experimentalists often rediscover it obliquely. For instance, some have worried that subjects may take a strategic attitude towards them – play against them as it were – because of the contrived experimental conditions, and thus may *not* express their preferences genuinely.⁷⁵ If this kind of attitude is to be excluded, another parallelism assumption must be added to the first, and to

decide whether or not it counts as an auxiliary hypothesis is again a matter of striking a balance between metaphysical and empirical claims.

The many reasons for underdetermination did not stop decision theorists (and at least some economists) from judging that VNM theory was empirically refuted, however differently they phrased this conclusion. Was it based on a two-sided comparison of the effects with the target theory, or did it appeal to alternative theoretical hypotheses? Here we meet an issue that attracted the attention of the philosophers of science when Lakatos's "methodology of scientific research programmes" (MSRP) made the existence of these hypotheses essential to the meaning of falsification. One of his claims is that a theory T counts as falsified *only if* there already exists an empirically progressive alternative T' within the same research programme. Lakatos makes this necessary condition *also sufficient*, and thus allows for the possibility that T be falsified simply because there exists a superior T' , i.e., even if there is no counterevidence.⁷⁶ Such a wedge between the senses of "falsification" and "refutation", the better understood term, makes the sufficiency claim dubious, so we leave it out and just discuss the more basic necessity claim.

The time order matters here, as it should in an historicized conception like the MSRP. Granting that VNM theory was generally taken to be refuted by the early 1980s, the natural candidates to the status of a superseding alternative theory are two-fold, i.e., Allais's long preexisting "positive theory" and Kahneman and Tversky's 1979 "prospect theory", and we have shown that both fall short of the desired status. We do not mean to say, of course, that unorthodox theoretical construals did not underlay the counterevidence, in the senses of suggesting it and helping to conceptualize it; but Lakatos's claim is more exacting than this point about theory-laddeness. As also explained, GEUT and "rank-dependent utility theory" (RDEUT) do meet the requirement for being theories, but only the latter is clearly an alternative, and historically they came after the fact. In sum, VNM theory exemplifies the case of a T that was taken to be refuted *without an alternative T' being yet in place*. This is a clear rebuttal of Lakatos's claim in its descriptive reading, and to say the least, a challenge to its normative reading, given the proclaimed endorsement of the "scientific elite" by the MSRP. By contrast, the conclusion is reassuring both for Duhem and Popper, who do not condition the occurrence of a refutation on Lakatos's heavy condition.⁷⁷

The time order is not all that matters in this discussion because the MSRP criterion for T' to count as a *superior* alternative would anyway not be met easily. This criterion requires in particular that T' have "excess empirical content" with respect to T , i.e., "predict novel facts", but that "all the unrefuted content of T [be] included", so that " T' explains the previous success of T " (Lakatos, 1970, p. 116).⁷⁸ Now, GEUT is mathematically stated as a logical weakening of VNM theory, which means that it cannot make predictions that the latter would not already make. It is only in the *psychological* sense that it has brought novel facts to attention – in principle, they could have been derived from the earlier theory.⁷⁹ RDEUT is in a somewhat different situation. It is similarly a weakening of EUT if one allows for the probability-transformation function to be the identity, but it is not anymore so with the commonly envisaged non-linear shapes, and it can then deliver new predictions compared with VNM theory. However, with RDEUT understood this way, the

other Lakatosian condition that T' recovers *all* unrefuted predictions of T becomes problematic. To summarize bluntly, the genuine alternative theories appear to be *sometimes too weak and sometimes too strong*.

Although progress is not the main topic of this paper, this finding should be related to the definition of this concept not only in Lakatos (1970, pp. 118–119), but also in Popper (1963–1972, pp. 240–243). The former took from the latter the grand view that science progresses by offering more informative theories that do not lose in “corroborated” content – the best of two worlds as it were. Duhem is too sceptical to make the heavy demands on progress of the two falsificationists, and his philosophy of science seems to be compatible with the view that RDEUT constitutes a genuine advance of decision theory with respect to VNM theory.⁸⁰

5.3 *Duhemian Confusion of the Sources of Falsity Within the Theory*

As we argued, Duhem’s underdetermination problem can sometimes be resolved to the point where the scientist holds a primary hypothesis responsible for the refutation, whence the question: can decision theorists say what exactly was refuted in VNM theory? The utility evaluation effect is telling against VNM theory in general, but involves too much of its content to be relevant at this stage. More to the point, the common consequence and common ratio effects are motivated by an attempt at questioning the VNM independence axiom (A3) specifically, and what they do achieve well is to *keep out of consideration the ordering and continuity axioms (A1) and (A2)*. Philosophers of science should praise those theorists, like Allais, who conceived of the discriminating experiments, and those experimentalists, like MacCrimmon and Larsson, or Kahneman and Tversky, who implemented them with relatively clear-cut results. Popper’s recommended procedure – i.e., once the theory is decomposed axiomatically, to devise tests for its axioms individually – appears to have been carried out with some success in the particular instance.

A comparison with earlier work will help to appreciate this success more fully. Mosteller and Nogee (1951) attempted to test VNM theory by using the strategy of estimating individual utility functions first. More precisely, they adapted the certainty equivalence method of estimation, which was described in 4.4, because they doubted that the individuals satisfied the transitivity condition in (A1) exactly. So they put the same questions several times over to their subjects and defined the monetary equivalent of a lottery for one of these subjects by some statistical criterion. Similarly, when they moved from the estimation to the testing stage, they repeated their questions and stated their results in terms of another statistic. Mosteller and Nogee’s conclude that “the notion that people behave in such a way as to maximize their expected utility is not unreasonable” (1951, p. 403), but even such a moderately favourable conclusion is far-fetched in view of the statistical assumptions that predetermine their data. Although this may be excused by the date, their work appears to be a sad example of mismanagement of Duhem’s underdetermination

problem. They *increased* the range of underdetermination instead of reducing it. Conceptually, their mistake was to involve (A1) in the analysis of the experiment, and it was aggravated by their way of replacing it not with a proper axiom, but with a compound of auxiliary hypotheses.

The generous interpretation of early experimental pieces like Mosteller and Nogee's is that they did not really aim at checking the axioms of VNM theory, being primarily concerned with lower-level propositions that this theory made it possible to formulate, such as those relative to risk attitudes or the subjective perception of probabilities. With this interpretation, it would become possible to claim that decision theorists were competent scientists throughout. In the first years, they would not have bungled the underdetermination problem, as we just suggested, but rather ignored it because they were working *within* the theory, and not yet curious about its axiomatic basis. The succession of a dogmatic and critical stage is standard in growth-of-knowledge accounts of science, like Kuhn's and Lakatos's, and although an oversimplification of our narrative, it appears to capture a grain of truth in it.

However we conclude on the first period, we must temper the praise of the second, because *it did not resolve the underdetermination problem fully*. The common consequence and common ratio effects hit (A3) and the reduction of compound lottery axiom jointly. Critics of VNM theory usually assumed that it was the former, not the latter, which was responsible for the effects, but one is struck by the meagre evidence they had at their disposal for this strong conclusion. Essentially, there was the point claimed by Allais, and to some extent borne out by experiments, that paradoxical choices were about as frequent among subjects who knew the probability calculus as among subjects who did not. This is hardly impressive because, as we mentioned, reduction can fail for more than one reason. Beside not knowing the multiplication rule of probabilities, an individual can make mistakes in applying it, and it is also conceivable that he masters the rule fully but doubts that probabilistic independence applies to the multi-stage lottery presented to him. Here, decision theorists are found lacking *on Duhemian terms*, which shows that the flexibility of the account is not without limit. There is an impressive convergence of the three effects, but it does not point towards (A3) alone, and more experiments were actually required.

It is indeed easy to test (A3) alone – select two pairs of lotteries with a common consequence that is not the same in each pair, regardless of the fact that some of the lotteries may be compound. Outside our period of study, this straightforward test was performed with the striking result that the proportion of violations was significantly weaker than that of the common consequence effect, and these violations proved to be irregular, arguably lacking the systematic element of the latter effect (see Conslik, 1989). An interesting view that has recently emerged is that the systematic violations of VNM theory must be referred to *independence in reduced form*, which is a joint implication of independence and reduction.⁸¹

Pending cross-confirmation of this hypothesis, we can say no more on it, but a point of philosophical interest has hopefully appeared in full light. *The solution to Duhem's underdetermination problem is sensitive to the axiomatic decomposition adopted for the theory*. If the responsibility of the decision-theoretic effects

eventually lies with a condition that is neither VNM independence, nor reduction of compound lotteries, the standard axiom system becomes unhelpful, and decision theorists had better replace it by a system – if it can be devised logically – that includes the culprit as one of its axioms. In terms of the standard system, there is only one sensible answer to underdetermination, which is – disappointingly – that independence and reduction are *jointly* responsible for the refutation. To the best of our knowledge, those who offer the axiomatic method as a remedy to underdetermination have not pointed out that the system can limit the degree to which the latter can be overcome.

5.4 *The Normative Discussion in the Quasi-Duhemian Account*

As they are usually understood, anomalies and refutations belong to the class of empirical counterexamples, the only ones thoroughly discussed in philosophy of science. But the VNM axioms were thought of as putative rationality conditions, and the theory was assessed in terms of normative force no less frequently than it was in terms of empirical performance. Sometimes the two viewpoints combined subtly, as in the Allaisian experiments about the wise men's rules of choice. A major question for our Duhemian perspective is whether it can make room for both the normative and empirical strands of the discussion.

The received history of VNM theory answers the problem in reorganizing the narrative roughly thus. Allegedly, the theory was offered as being primarily empirical. But then came Allais with his paradoxes, which made it clear that it was descriptively false. The tenors of EUT had to recognize the fact, especially Savage, whose own answers to Allais's questionnaire documented the failure, and what they did afterwards was to *retreat to the normative*. Soon followed by others, Savage, de Finetti, Marschak and Samuelson buttressed VNM theory by expanding on its normative justifications. They left the empirical ground to the experimentalists, who developed the new "theories" on this ground more or less exclusively.

The historical part of the paper contradicts this reconstruction in a number of ways. First, VNM theory was proposed as a theory of rational choice right from the beginning – this is how it appears in *Theory of Games and Individual Behaviour* as well as in Marschak (1950). Second, Allais's paradoxes were thought experiments specially devised to challenge the claim that VNM theory was normatively compelling, and nobody took them to constitute empirical refutations until they were transformed by experimentalists much later. Third, even at this late stage, defenders of VNM theory had empirical arguments at their disposal and sometimes made use of them. The only indisputable claim is that Samuelson and Savage neglected these arguments, but they had hardly paid attention to the empirical side of the theory at first.⁸²

In sum, the two components were present at all temporal stages of the discussion, and often, though admittedly not always, in the one and the same piece of work. This suggests focusing on *arguments*, rather than times or individuals, as the proper

units of analysis. Here is a very brief review of those which can be expected to link the normative and the empirical with each other.

It is at least arguable that *a rational rule of conduct stands a better chance of being adopted than an irrational one*. There is some evidence that the VNM theorists conceived of this linkage, and by this plausible attribution, one makes sense of the claim that the normative defences – especially, those mentioned in Section 3 – influenced the acceptance of von Neumann and Morgenstern’s theory as an empirical one.⁸³ The linkage of rationality with frequency can also be presented in the converse way: gross violations of rationality are not met very often. This version surfaced in some of the controversies about the paradoxes, and up to now, it remains the best reason that decision theorists have at their disposal to impose the dominance principle on their empirical theories.

Another connecting claim is that *strict rationality conditions give structure to the choice data*, whether these conditions are themselves put to the test (like VNM independence) or assumed to hold in order to facilitate the test of other conditions (like transitivity). This linkage is unproblematic, contrary to the previous one, but it may be unspecific to the theories of rational decision making, since it is always the case that a hypothesis easier to test the logically stronger it is.

Still another connection is that *the normative force of a rationality condition may be subjected to a test*, as in the argument about Allais’s paradox. Remember that Allais was initially concerned with experimentation only to check that rational and prudent people made the choice that he had foreshadowed in his thought experiment. He avoided crude circularity by assuming that rationality and prudence are qualities that can be recognized in someone by common consent. Savage also relied on an empirical test of normativity, though a different one. He argued in effect that if a man decides in a rational moment not to comply with a rationality condition, this establishes that the condition has little or no force after all. He avoided crude circularity by assuming that a rational moment can be recognized at the time and depth of thinking given to the issue.

We finally mention the connection implied by *the cognitive preconditions of rational decision-making*. Since Simon’s work on bounded rationality, it is a well taken point that in order to have any normative force, a rule of decision must make feasible demands on the individual’s ability to collect information and make computations. This new linkage of the normative and the empirical is but a contextual way of making good the meta-ethical principle that “ought” implies “can”. We found it mentioned by MacCrimmon and Larsson at the expense of VNM theory, but it can be no less damaging against alternatives.

How does this sketchy and no doubt incomplete list reflect on the quasi-Duhemian account? Without providing the full argument, we submit that it does not undermine it, but rather serves to determine it further. Some of the suggested links between the normative and the empirical are expressed in claims that can *themselves* be tested, whether directly or indirectly, thus delivering a new range of possibilities for Duhemian underdetermination. One such claim is that gross violations of rationality cannot be met very often. A possible check is to observe whether or not actual decisions obey the dominance principle, a somewhat neglected test. If the answer is

significantly negative, there will be a – Duhemian – choice between giving up the claim or disconnecting the dominance principle from rationality. Other links point in the direction of probing the normative force of an axiom in the way exemplified by MacCrimmon and Larsson; we have dealt with this case at some length. Still other links correspond to *higher-order claims constraining the interpretation of tests*. For instance, the cognitive version of the principle that “ought” implies “can” may be invoked to blame the compound lottery axiom, rather than VNM independence, for the common consequence and common ratio effects. To handle this more disturbing category, but we suggest adding the claims to the considerations that make it possible for decision theorists to resolve the confusion of hypotheses. Here we take up Duhem’s “bon sens” while interpreting it very liberally; we boldly enlarge it to include the metaphysical and heuristic commitments that regulate the functioning of a scientific discipline. Because Duhem does not clearly say that universal Cartesian commonsense is not sufficient, and that physics would not arbitrate its internal conflicts without preconceptions of this more specific sort, our account is not exclusively moulded after him, but is only quasi-Duhemian. A whiff of Lakatosian MSRP proves to be necessary after all.⁸⁴

Acknowledgement The author is grateful to Mohammed Abdellaoui for technical discussions on decision theory and to Anastasios Brenner and Brian Hill for philosophical and editorial comments.

Endnotes

- 1 Harding’s collection (1976) is the locus classicus for the so-called Duhem-Quine thesis. See also Lakatos (1970, p. 180 sq). Among those who have criticized the conflation are Vuillemin (1979), Ariew (1984), Boyer (1994), Gillies (1993, Chap. 5).
- 2 See chaps. IV and V in the second part of *TP*.
- 3 See Fodor and Le Pore (1992) for relevant clarification. Note that the distinction between two forms of holism does not coincide with the historical difference between Duhem and Quine, because the latter, if perhaps not the former, promoted both forms.
- 4 *TP*, II, VI, §I, pp. 273–278; English transl. pp. 180–183.
- 5 Note the further contrast between Duhem and Quine: the latter would assuredly not restrict his underdetermination thesis to a particular group of sciences. Logically, this point is distinct from the already mentioned one that Quine extends the degree of underdetermination beyond what Duhem wishes, but the two points are entangled in Quine’s exposition (e.g., *FLPV*, p. 42).
- 6 We will concentrate on falsificationism because this school has more interest in Duhem’s underdetermination thesis than any other, and has made bold attempts at offering a general solution to the problem it raises. Bayesian philosophy does not seem to us to score any better, nor any alternative attempt that is currently available, but this further argument is not for this paper.
- 7 We refer to Brenner (1990) for a full discussion of how Duhem integrated the history and philosophy of science with each other.
- 8 Those of Machina (1983, 1987), Sugden (1986) and Munier (1988) still make good reading. Camerer (1995), Cohen and Tallon (2000), and Starmer (2000) extend them thoroughly, while complementing each other well.

- 9 Mongin (1997) gives an English summary of this French paper. Cross (1982) has precedence for discussing the underdetermination problem in economics, with privileged examples from macroeconomics and macroeconometrics. More papers have followed, among which Sawyer, Bid and Sankey's (1997), and they are still generally oriented towards macro-applications. Note that all this literature relies on the construal of a "Duhem-Quine thesis".
- 10 The title of Harding's (1974) collection, *Can Theories Be Refuted?*, is by itself indicative of the trend. But oddly enough, most of the essays skip the connection between Duhem and Popper. Among the recent textbooks in philosophy of science, Gillies's (1993) is exceptional in mentioning this connection.
- 11 On Wiener's experiment, see *TP*, II, VI, §II, pp. 279–282 (English transl., pp. 184–186) and especially the items listed at the end of this passage. Foucault's experiment is discussed for the first time on pp. 282–283 (English transl. pp. 186–187), and Duhem returns to it several times, so it may be viewed as paradigmatic for his underdetermination thesis.
- 12 See, e.g., Grünbaum's (1960) and Laudan's (1965) papers reprinted in Harding's (1976) collection.
- 13 The exception is Grünbaum (1960), who irrelevantly attributed (a') on top of (a) to Duhem. Laudan (1965) corrected him, and in doing this, proposed the now received distinction between a weak and a strong form of the "Duhem-Quine thesis" (in Harding, 1974, p. 159).
- 14 Grünbaum's (1960) recognizes that he cannot provide a criterion beyond the exclusion of the most obvious form of circularity (in Harding, 1974, pp. 181).
- 15 Actually, despite the popular "come what may", even Quine does not seem to entertain the strong form seriously.
- 16 See, e.g., Lakatos (1970, p. 184). He refers to the statement of the thesis in terms of rationality as to the "strong interpretation", which is confusing given the preexisting distinction between a "weak" and "strong" form of the thesis.
- 17 Mongin (1988) made a similar distinction between "thèse" and "problème de Duhem".
- 18 For a Bayesian analysis of test and confirmation, see Howson and Urbach (1993, Chap. 7).
- 19 This is the maximal amplification of Duhem's epistemological holism; it appears in *TP*, II, VI, §II, p. 285 and §V, pp. 303–304; English transl., pp. 187–188, 199–200. Brenner (1990, p. 226) suggests connecting it with Duhem's usual emphasis on the theoretical unity of physics.
- 20 See Popper (1935–1972, § 19, n. 1 and § 85, n. *2; 1963–1972, pp. 112–113).
- 21 The argument is also in Duhem (1908, pp. 132–133).
- 22 See, e.g., Laugier (1999).
- 23 Both lines of reasoning can indeed be found in Popper (1963–1972, pp. 238–239).
- 24 See, e.g.: "Our new theory will represent a potential step forward, whatever the outcome of the new tests may be. For it will be better testable than the previous theory" (1963–1972, p. 242).
- 25 See Lakatos (1970, pp. 116–120). This passage defines what it means for a theory to be properly "falsified" and for a series of theories to be "progressive". Lakatos is clearly struggling here with "the Duhem-Quine thesis" even if he does not mention it.
- 26 Mongin (2003) discusses the sense in which this style of axiomatization meets the requirements that logicians usually put on the axiomatic method.
- 27 See the proceedings, pp. 143 and 163, as well as Samuelson (1952a). For more detail on the origin of the independence condition, see Fishburn (1989) and Fishburn and Wakker (1995).
- 28 We skip the technical formulation. There are various definitions of continuity, the choice depending on how VNM independence is simultaneously defined. For a mathematical comparison between the systems, see Fishburn (1982).
- 29 Compare Luce and Raiffa's approach to compound lotteries with Samuelson's (1952b, pp. 671). Starting from the same definition in terms of prizes, he assumes reduction of compound lotteries to hold as a "convention", not as substantial and possibly behavioral condition.
- 30 It is another curious slip that the preliminary step based on (A1) and (A2) was fully clarified after the VNM representation theorem had been proved and become widely known.

- 31 This connection goes back to Marshall's *Principles of Economics* (1890–1921) through von Neumann and Morgenstern's *Theory of Games* (1944–1947, p. 28, 629). For a recent analysis of the pleasure of game, see Diecidue, Schmidt and Wakker (2004).
- 32 Later, Ellsberg (1961) came to question Savage's (1954) axioms by means of a thought-provoking experiment, and he is remembered today mostly for this contribution.
- 33 For a fuller criticism, see McClennen (1983), to whom we also refer for an appraisal of the next justification.
- 34 See in particular Hammond (1988) and Machina (1991). The first argument along Samuelson's line is to be found, even with more detail, in an unpublished paper by Rubin (1949).
- 35 See Marshall (1890–1921, pp. 111–112, taken up in *Mathematical Appendix, Note IX*, pp. 693–694).
- 36 The hypothesis was also disputed by Markowitz (1952) on the theoretical ground that it did not properly distinguish between the individuals' current income and total wealth when a lottery is drawn out.
- 37 Some simple applications can be found in today's microeconomics texts like Varian's (1978).
- 38 "Fondements et applications de la théorie du risque en économétrie", organized by Centre National de la Recherche Scientifique, 12–17 May 1952.
- 39 Savage's *Foundations of Statistics* appeared only in 1954, but he presented some of its content in Paris, and this was hailed as an advance compared with VNM theory even by opponents to the latter, like Allais.
- 40 There is a two-choice format of experiment adapted to Savage's system, and especially to the analogue in it of the VNM independence condition. This format became established in 1961 with the so-called Ellsberg paradox, which involves choosing between urns instead of lotteries.
- 41 See Jallais and Pradier's (2005) investigation of when and how the Allais paradox emerged in 1952.
- 42 The proceedings of the 1952 conference contain some applications of the two-pair format, but they are due to two other French engineers, Massé and Morlat. Their contribution anticipates on Allais's critique of EUT significantly.
- 43 Mongin (1988) fell into the trap. Guala (2000) argued against him that normative issues had decisively influenced both Allais's conception and the others' reception of the paradox. The present account corrects the bias in the earlier one.
- 44 As early as 1952, he began a vast questionnaire study, but the answers proved to be difficult to exploit, and only some results were published as late as 1979. See Allais (1979, pp. 447–448) for details.
- 45 See Allais (1953, p. 518). We leave out a third component, which is the use of objective probabilities, because Allais seems to vary about its rôle and meaning.
- 46 At the 1952 conference, de Finetti offered it as the major argument for EUT; see the proceedings, pp. 159, 196. It interacts in a complex way with the issue of the "cardinality" of the VNM function and the $u(x)$ that replaces it in Allais's positive conception. We had to leave this major topic aside from the account to keep it tractable, and instead refer to Allais's (1953a, b, 1979) discussion, along with Fishburn's (1989) and Bouyssou and Vansnick's (1990) clarifying accounts.
- 47 This is fallback line. Apparently, Savage was not prepared anymore to make the strong normative claims of his 1952 paper with Friedman.
- 48 For more details, see the survey part of MacCrimmon and Larsson (1979, pp. 364–366). The experiment involving reconsideration was made by Slovic and Tversky (1974).
- 49 Neither of these comments is made by MacCrimmon and Larsson.
- 50 However, part of Allais's suppositions on the wise men remain untested. MacCrimmon and Larsson did not control for the subjects' acquaintance with the probability calculus, and in particular, for their acceptance of the compound lottery axiom.
- 51 Specifically, method 1, which proposes first a lottery $l = ((0.98, M), (0.02, 0))$, where M is the maximum amount of money under consideration, and then $l' = ((0.5, l), (0.5, 0)) =$

- ((0.49,M),(0.51,0)) may not lead to the same result as Method 2, which starts from I' right away.
- 52 For more on the utility evaluation effect, see the surveys by Machina (1983, 1987) and Jaf-ray (1989).
 - 53 Hey (1991) and Camerer (1995) cover more ground. Our selection is constrained by the philo-sophical purpose and fixed time limits, and for expository simplicity, we left out the works, including Loomes and Sugden's (1982) and Fishburn's (1988), which relax the ordering axiom at the same time as VNM independence.
 - 54 See the second paper at the Paris conference (pp. 127–140 of the 1953 proceedings).
 - 55 Allais (1953b, pp. 511–513). This accords with his comments after Samuelson's and Massé and Morlat's papers (pp. 154 and 194–195 of the 1953 proceedings).
 - 56 See Allais (1979).
 - 57 Hagen's formula has another possible interpretation in terms of disappointment, which Loomes and Sugden (1986) stress, thus connecting it with a model of their own.
 - 58 References are to be found in Quiggin (1982) and Camerer (1995).
 - 59 For the first point: the authors themselves revised "prospect theory" in the direction of RDEUT, which led to "cumulative prospect theory" (Tversky and Kahneman, 1992). For the second point: Karmakar (1978) independently put forward a S-shaped form for a to accom-modate the utility evaluation effect as well as the Allais paradox, and this shape seems to have received no less experimental support than the convex shape.
 - 60 Technically, it is just a convex subset of a vector space, not itself a vector space.
 - 61 Beside claiming qualitative association of risk-aversion with concavity, and of risk-love with convexity, VNM theory uses the "Arrow-Pratt index of risk-aversion" as a quantitative meas-ure (see, e.g., Varian, 1978 for an elementary exposition). Machina's generalizations are in particular concerned with retaining the properties of this index.
 - 62 Two dimensions are sufficient to represent preferences over three-outcome lotteries because of the normalization $p_1 + p_2 + p_3 = 1$. This geometric representation was introduced by Marschak (1950) to liken VNM theory with the neo-classical theories of choice under certainty.
 - 63 A stronger, but perhaps not very plausible interpretation, is that for any given pair of choices that violates VNM theory, if one assumes (H1), all local utility functions that are compatible with the violation turn out to satisfy (H2).
 - 64 Despite being weak in the sense explained, (H2) may be too strong in another sense, because it does not accommodate all replications of the common consequence or common ratio effects. Camerer (1995, p. 636) suggests that it should applied to only part of the set L .
 - 65 Taking the particular case $u(x) = x$, Yaari (1987, p. 107) demonstrates that the f function is convex throughout the 0–1 interval if and only if the individual is risk averse in the sense of preferring any lottery I to a variant of I in which the outcomes are statistically more dispersed. For a more general theorem, see Chew, Karni and Safra (1987). Cohen and Tallon (2000) survey these results.
 - 66 By adopting a treatment of risk-attitudes in terms of f , they become able to use u for other semantic purposes, in particular to recover the time-honoured neo-classical hypothesis of decreasing marginal satisfaction.
 - 67 See the sketch in Yaari (1987, p. 104).
 - 68 See also Gilboa (1987) and Schmeidler (1989).
 - 69 See again Camerer's assessment (1995).
 - 70 Mongin (1988, p. 312) did not separate the two interpretations sufficiently well. Granger (1992, p. 244) suggested that the latter was more appropriate than the former at this stage of the empirical discussion, and we are following him now.
 - 71 Compare with, e.g., the textbooks by Hey (1991) and Friedman and Sunder (1994).
 - 72 See Guala (2005, Chap. 7) for further discussion of the parallelism issue.
 - 73 Hausman and Mongin (1998) evidence and discuss this kind of dogmatism.
 - 74 Furthermore, some would add the condition that choice data be amenable to standard econo-metric methods. The experimental work discussed in this paper relies on straightforward

- counting of proportions, but the field has witnessed increased use of econometrics since the 1990s.
- 75 An analogy here is that of the elector who votes for a non-preferred candidate because of the electoral system.
- 76 A scientific theory T is falsified if and only if another theory T' has been proposed with the following characteristics: (1) T' has excess empirical content over T ...; (2) T' explains the previous success of T ...; (3) some of the excess content of T' is unrefuted" (Lakatos, 1970, p. 116, our emphasis).
- 77 The textual evidence for Duhem is the already mentioned analysis of Wiener's experiment. As reported in *TP* and earlier work, it was intended to refute Neumann's hypothesis and succeeded in doing so; no alternative was involved. That Popper allows for "falsifications" without superseding theories is clear from passages such as those already discussed on "crucial arguments" in *LSD* and *Conjectures and Refutations*.
- 78 Our discussion does not need Lakatos's last requirement that some of the novel predictions be "corroborated".
- 79 Machina's (1983, pp. 287–289) example fits this description.
- 80 Mongin (1988, 1997) already argued that this demand was inapplicable to the case. The present account sharpens the critique of Popper and Lakatos, while remaining within the broad confines of refutationism.
- 81 See Segal (1995) and Camerer (1995) for more on this interpretation.
- 82 A fourth problematic claim, which we cannot address here, is that the new construals were developed in a normative vacuum.
- 83 Concerning Samuelson's defence, Machina writes: "although this is a prescriptive argument, it has played a key role in economists' adoption of expected utility as a descriptive theory of choice under uncertainty" (1987, p. 127). The same could be said of Friedman and Savage's defence, despite its being flawed.
- 84 Compare with the limited use of MSRP in Mongin (2002) to tackle the issue of progress in normative theories.

Bibliography

- Fondements et applications de la théorie du risque en économétrie*, Paris, Centre National de la Recherche Scientifique, 1953.
- Allais Maurice (1953a), "Fondements d'une théorie positive des choix comportant un risque et critique des postulats et axiomes de l'école américaine", Appendix to *Fondements*, pp. 257–332. Reprint, Paris, Imprimerie Nationale, 1955. English transl. "The Foundations of a Positive Theory of Choice Involving Risk and a Criticism of the Postulates and Axioms of the American School", in Allais and Hagen, 1979, pp. 27–145.
- Allais Maurice (1953b), "Le comportement de l'homme rationnel devant le risque", *Econometrica*, 21, pp. 503–546.
- Allais Maurice (1979), "The So-called Allais Paradox and Rational Decisions Under Uncertainty", in Allais and Hagen, 1979, pp. 437–683.
- Allais Maurice (1988), "The General Theory of Random Choices in Relation to the Invariant Cardinal Utility Function and the Specific Probability Function. The (U, θ) Model: A General Overview", in Munier, 1988, pp. 231–289.
- Allais Maurice and Hagen O. (1979) (eds), *Expected Utility Hypothesis and the Allais Paradox*, Dordrecht, D. Reidel.
- Ariew Roger (1974), "The Duhem Thesis", *British Journal for the Philosophy of Science*, 35, pp. 313–325.
- Bouyssou Denis and Vansnick Jean-Claude (1990), "Utilité cardinale dans le certain et choix dans le risque", *Revue économique*, 6, pp. 979–1000.

- Boyer Alain (1994), *Introduction à la lecture de Karl Popper*, Paris, Presses de l'École normale supérieure.
- Brenner Anastasios (1990), *Duhem, science, réalité et apparence*, Paris, Vrin.
- Camerer Colin (1992), "Recent Tests of Generalizations of Expected Utility Theory", in Edwards, 1992, pp. 207–251.
- Camerer Colin (1995), "Individual Decision Making", in J.H. Kagel and A. Roth (eds), *The Handbook of Experimental Economics*, Princeton, Princeton University Press, Chap. 8, pp. 587–703.
- Chew Soo Hong, Karni Edi and Safra Zvi (1987), "Risk Aversion in the Theory of Expected Utility with Rank Dependent Probabilities", *Journal of Economic Theory*, 42, pp. 370–381.
- Cohen Michèle and Tallon Jean-Marc (2000), "Décision dans le risque et l'incertain: l'apport des modèles non-additifs", *Revue d'économie politique*, 110, pp. 631–681.
- Conslis John (1989), "Three Variants on the Allais Example", *American Economic Review*, 79, pp. 392–407.
- Cross Rod (1982), "The Duhem-Quine Thesis, Lakatos and the Appraisal of Theories in Macroeconomics", *Economic Journal*, 92, pp. 320–340.
- Diecidue Enrico, Schmidt Ulrich and Wakker Peter (2004), "The Utility of Gambling Reconsidered", *Journal of Risk and Uncertainty*, 29, pp. 241–259.
- Duhem Pierre (1906), *La théorie physique. Son objet, sa structure*, Paris, Chevalier et Rivière (2nd ed., 1914). Reprint, Paris, Vrin, 1981, intr. P. Brouzeng. *The Aim and Structure of Physical Theory*, English transl. by P. Wiener, Princeton, Princeton University Press, 1954.
- Duhem Pierre (1908), *Sozein ta phainomena. Essai sur la notion de théorie physique de Platon à Galilée*, Paris, Hermann. Reprint Paris, Vrin, 1994, intr. by P. Brouzeng. *To Save the Phenomena*, English transl. by E. Dolland and C. Maschler, Chicago, The University of Chicago Press, 1969.
- Edwards Ward (eds), (1992), *Utility Theories: Measurement and Applications*, Dordrecht, Kluwer.
- Ellsberg Daniel (1956), "Classic and Current Notions of 'Measurable Utility'", *Economic Journal*, 64: 528–556.
- Ellsberg Daniel (1961), "Risk, Ambiguity, and the Savage Axioms", *Quarterly Journal of Economics*, 75, pp. 643–669.
- Fishburn Peter C. (1970), *Utility Theory for Decision Making*, New York, Wiley.
- Fishburn Peter C. (1982), *The Foundations of Expected Utility*, Dordrecht, Reidel.
- Fishburn Peter C. (1988), *Nonlinear Preference and Utility Theory*, Baltimore, Johns Hopkins University Press.
- Fishburn Peter C. (1989), "Retrospective on the Utility Theory of von Neumann and Morgenstern", *Journal of Risk and Uncertainty*, 2, 127–158.
- Fishburn Peter C. and Wakker Peter (1995), "The Invention of the Independence Condition for Preferences", *Management Science*, 41, 1130–1144.
- Fodor Jerry and Lepore Ernest (1992), *Holism: A Shopper's Guide*, Oxford, Blackwell.
- Friedman Daniel and Sunder Shyam (1994), *Experimental Methods. A Primer for Economists*, Cambridge, Cambridge University Press.
- Friedman Milton and Savage Leonard J. (1948), "The Utility Analysis of Choices Involving Risk", *Journal of Political Economy*, 56, pp. 278–304.
- Friedman Milton and Savage Leonard J. (1952), "The Expected Utility Hypothesis and the Measurability of Utility", *Journal of Political Economy*, 60, pp. 463–474.
- Gilboa Itzhak (1987), "Expected Utility with Purely Subjective Non-Additive Probabilities", *Journal of Mathematical Economics*, 16, pp. 65–88.
- Gillies Donald (1993), *Philosophy of Science in the Twentieth Century*, Oxford, Blackwell.
- Granger Gilles (1992), *La vérification*, Paris, Odile Jacob.
- Grünbaum A. (1960), "The Duhemian Argument", *Philosophy of Science*, 27, 75–88, reprinted in Harding, 1974, pp. 116–131.
- Guala Francesco (2000), "The Logic of Normative Falsification: Rationality and Experiments in Decision Theory", *Journal of Economic Methodology*, 7, pp. 59–93.

- Guala Francesco (2005), *The Methodology of Experimental Economics*, Cambridge, Cambridge University Press.
- Hagen Ole (1979), "Towards a Positive Theory of Preferences under Risk", in Allais and Hagen, 1979, pp. 271–302.
- Hammond Peter J. (1988), "Consequentialist Foundations for Expected Utility", *Theory and Decision*, 25, 25–78.
- Hausman Daniel and Mongin Philippe (1998), "Economists' Responses to Anomalies: Full-Cost Pricing versus Preference Reversals", in J. Davis (eds), *New Economics and Its History. History of Political Economy, Annual Supplement*, vol. 29, pp. 255–272.
- Hey John D. (1991), *Experiments in Economics*, Oxford, Blackwell.
- Harding Sandra G. (eds) (1976), *Can Theories be Refuted? Essays on the Duhem-Quine Thesis*, Dordrecht, Reidel.
- Jaffray Jean-Yves (1989), "Some Experimental Findings on Decision-Making under Risk and their Implications", *European Journal of Operation Research*, 38, pp. 301–306.
- Jallais Sophie and Pradier Pierre-Charles (2005), "The Allais Paradox and its Immediate Consequences for Expected Utility Theory", in P. Fontaine and R. Leonard (eds), *The 'Experiment' in the History of Economics*, Chap. 4, London, Routledge.
- Kahneman Daniel and Tversky Amos (1979), "Prospect Theory: An Analysis of Decision under Risk", *Econometrica*, 47, pp. 263–291.
- Karmarkar Uday (1974), "The Effect of Probabilities on the Subjective Evaluation of Lotteries", WP no 698-74, MIT, Sloane School of Management.
- Karmarkar Uday (1978), "Subjective Weighted Utility: A Descriptive Extension of the Expected Utility Model", *Organizational Behavior and Human Performance*, 21, pp. 61–72.
- Lakatos Imre (1970), "The Methodology of Scientific Research Programmes", in I. Lakatos and A. Musgrave (eds), *Criticism and the Growth of Knowledge*, Cambridge University Press, pp. 91–196.
- Laudan Laurens. (1965), "Grünbaum on the Duhemian Argument", *Philosophy of Science*, 32, 295–301, reprinted in Harding, 1974, pp. 155–161.
- Laugier Sandra (1999), "Expérience cruciale", in D. Lecourt (eds), *Dictionnaire d'histoire et philosophie des sciences*, pp. 404–406.
- Loomes Graham and Sugden Robert (1982), "Regret Theory: An Alternative Theory of Rational Choice under Uncertainty", *Economic Journal*, 92, pp. 805–824.
- Loomes Graham and Sugden Robert (1986), "Disappointment and Dynamic Consistency in Choice under Uncertainty", *Review of Economic Studies*, 53, pp. 271–282.
- Luce R. Duncan and Raiffa Howard (1957), *Games and Decisions*, New York, Wiley, 1957.
- McCord Mark and Neufville Richard de (1983), "Empirical Demonstration That Expected Utility Is Not Operational", in Stigum and Wenstop, 1983, pp. 181–199.
- McClelland Edward F. (1983), "Sure-Thing Doubts", in Stigum and Wenstop, 1983, pp. 117–136.
- MacCrimmon Kenneth R. and Larsson Stig (1979), "Utility Theory: Axioms versus 'Paradoxes'", in Allais and Hagen, 1979, pp. 333–409.
- Machina Mark J. (1982), "'Expected Utility' Analysis without the Independence Axiom", *Econometrica*, 50, pp. 277–323.
- Machina Mark J. (1983), "Generalized Expected Utility Analysis and the Nature of Observed Violations of the Independence Axiom", in Stigum and Wenstop (1983, pp. 263–293).
- Machina Mark J. (1987), "Choice under Uncertainty: Problems Solved and Unsolved", *Journal of Economic Perspectives*, 1, pp. 121–154.
- Machina Mark J. (1991), "Dynamic Consistency and Non-Expected Utility", in M. Bacharach and S. Hurley (eds), *Foundations of Decision Theory*, Oxford, Blackwell, pp. 39–91.
- Markowitz Harry (1952), "The Utility of Wealth", *Journal of Political Economy*, 60, pp. 151–158.
- Marshall Alfred (1890), *Principles of Economics*, London, MacMillan, 8th ed., 1920.
- Marschak Jacob (1950), Rational Behavior, "Uncertain Prospects, and Measurable Utility", *Econometrica*, 18, pp. 111–141.

- Mongin Philippe (1988), "Problèmes de Duhem en théorie de l'utilité espérée", *Fundamenta Scientiae*, 9, pp. 289–317.
- Mongin Philippe (1997), "Expected Utility Theory", in J. Davis, W. Hands and U. Mäki (eds), *Handbook of Economic Methodology*, London, Elgar, 1997, pp. 342–350.
- Mongin Philippe (2003), "L'axiomatisation et les théories économiques", *Revue économique*, 54, pp. 99–103.
- Mongin Philippe (2006), "A Concept of Progress for Normative Economics", *Economics and Philosophy*, 22, pp. 19–54.
- Mosteller Frederick and Nogee Philip (1951), "An Experimental Measurement of Utility", *Journal of Political Economy*, 59, 371–404.
- Munier Bertrand (1988), "A Guide to Decision-Making under Uncertainty", in Munier, 1988, pp. 1–34.
- Munier Bertrand (eds), (1988), *Risk, Decision, and Rationality*, Dordrecht, Reidel.
- Popper Karl (1935), *Logik der Forschung*, Vienna, Julius Springer. Transl. as *The Logic of Scientific Discovery*, London, Hutchinson, 1959 (6th revised ed., 1972).
- Popper Karl (1963), *Conjectures and Refutations*, London, Routledge (4th revised ed., 1972).
- Popper Karl (1983), *The Postscript to the Logic of Scientific Discovery*, I. Realism and the Aim of Science, London, Hutchinson.
- Quiggin John (1982), "A Theory of Anticipated Utility", *Journal of Economic Behavior and Organizations*, 3, 323–343.
- Quine Willard V.O. (1953), *From a Logical Point of View*, Cambridge, Mass., Harvard University Press, 1980.
- Rubin Herman (1949), "The Existence of Measurable Utility and Psychological Probability", *Cowles Discussion Papers*, no 332.
- Samuelson, Paul A. (1952a), "Utility, Preference, and Probability" in J. Stiglitz (eds), *The Collected Scientific Papers of Paul A. Samuelson*, 1, Chap. 13, 127–136.
- Samuelson Paul A. (1952b), "Probability, Utility, and the Independence Axiom", *Econometric Society*, 20, 670–678.
- Savage Leonard J. (1954), *The Foundations of Statistics*, New York, Wiley, 2nd ed., New York, Dover, 1972.
- Schmeidler David (1986), "Integral Representation without Additivity", *Proceedings of the American Mathematical Society*, 97, 255–261.
- Schmeidler David (1989), "Subjective Probability and Expected Utility without Additivity", *Econometrica*, 57, 571–587.
- Segal Uzi (1992), "The Independence Axiom versus the Reduction Axiom: Must We Have Both?", in Edwards, p. 165–183.
- Slovic Paul and Tversky Amos (1974), "Who Accepts Savage's Axiom?", *Behavioral Science*, 19, 368–373.
- Starmer Chris (2000), "Developments in Non-expected Utility Theory: the Hunt for a Descriptive Theory of Choice under Risk", *Journal of Economic Literature*, 38, 332–382.
- Sugden Robert (1986), "New Developments in the Theory of Choice Under Uncertainty", *Bulletin of Economic Research*, 38, 1–24.
- Stigum Bernt and Wenstop Fred (1983) (eds), *Foundations of Utility and Risk theory with Applications*, Dordrecht, Reidel.
- Varian Hal R. (1978), *Microeconomic Analysis*, New York, Norton.
- Von Neumann John and Morgenstern Oskar (1944), *The Theory of Games and Economic Behavior*, Princeton, Princeton University Press (2nd ed. 1947).
- Vuillemin Jules (1979), "On Duhem and Quine's Theses", *Grazer Philosophische Studien*, 9, 69–96.
- Yaari Menahem (1987), "The Dual Theory of Choice under Risk", *Econometrica*, 55, 95–115.

Notes on the Authors

Daniel Andler was trained in mathematics and philosophy in Paris and at UC Berkeley. A specialist of model theory, he first taught mathematics, then moved to positions in philosophy. He now teaches at Université Paris-Sorbonne (Paris IV). He recently stepped down as founding head of the Department of Cognitive Studies at École normale supérieure. He is the current President of the Société de philosophie des sciences. In 2007 he was appointed to the Institut universitaire de France. His central interest lies in the foundations of cognitive science, which he approaches from the dual perspective of philosophy of science and philosophical anthropology. He works on specific issues concerning models of the mind, the role of context, and reasoning. He defends a minimal version of naturalism, with the aim of articulating cognitive science with the social sciences. His publications include: “The Normativity of Context”, *Philosophical Studies*, 100, 2000; (co-author) *Philosophie des sciences*, Gallimard, 2002; “Context: the Case for a Principled Epistemic Particularism”, *Journal of Pragmatics*, 35–3, 2003; (editor & co-author) *Introduction aux sciences cognitives*, Gallimard, 1992, augm. ed. 2004; “Phenomenology in Cognitive Science and Artificial Intelligence”, in *A Companion to Phenomenology and Existentialism*, Blackwell, 2006.

Contact: Département d'études cognitives, École normale supérieure, 29 rue d'Ulm, F-75005 Paris. daniel.andler@ens.fr

Bernadette Bensaude-Vincent got a PhD in Philosophy at Paris-Sorbonne (Michel Serres supervisor). She then became a member of the Research Center in History of Science headed by Robert Fox at the Cité des sciences et de l'industrie. She is now a Professor of History and Philosophy of Science at the Université Paris X. Her research interests focus on the history and philosophy of chemistry. She is presently conducting a research program on nanotechnology: ethical and philosophical issues. She is a member of the Académie des technologies and of the Ethics Committee of the Centre national de la recherche scientifique. Among her publications: *Lavoisier, mémoires d'une révolution* (1993); *A History of Chemistry* (with Isabelle Stengers, 1996); *Eloge du mixte: Matériaux nouveaux et philosophie ancienne* (1998); *Se libérer de la matière? Fantômes autour des nouvelles technologies* (2004); *Chemistry, The Impure Science* (with J. Simon) forthcoming.

Contact: Département de philosophie, Université Paris X, 200 avenue de la République, F-92001 Nanterre. bensaude@u-paris10.fr

Anastasios Brenner is professor of philosophy at the Université Paul Valéry (Montpellier). Born in New York City in 1959, he studied philosophy at the Université de Paris-Sorbonne. His doctorate was on the relation of philosophy of science and history of science in the work of Duhem (*Duhem: Science, réalité et apparence*, Paris, Vrin, 1990). After editing Duhem, *L'aube du savoir: Épitomé du Système du monde* (Paris, Hermann, 1997), he focused on Poincaré, the conventionalist movement and its impact on the Vienna Circle (*Les origines françaises de la philosophie des sciences*, Paris, PUF, 2003). Recent articles are “Quelle épistémologie historique? Kuhn, Feyerabend, Hacking et l'école Bachelardienne” (*Revue de métaphysique et de morale*, 1, 2006); “Classification des sciences et encyclopédie: Neurath et la tradition française” (T. Martin ed., *Le tout et les parties dans les systèmes naturels*, Paris, Vuibert, 2007). His main areas of interest are the historical development of philosophy of science, the relation between contemporary philosophy and science and the rational criteria involved in theory choice. He is a founding member and vice-president of the Société de philosophie des sciences and chaired the organizing committee of the Sixth International History of Philosophy of Science Congress (HOPOS 2006) held in Paris.

Contact: Département de philosophie, Université Paul Valéry, route de Mende, F-34199 Montpellier cedex 5. anastasios.brenner@wanadoo.fr

Gilles Cohen-Tannoudji, born in Constantine (Algeria) in 1938, is a graduate of the École Polytechnique (1958–1960). He has done his whole career in the French Atomic Energy Commission (CEA) at Saclay, from which he retired in 1998. He is currently emeritus researcher in the Laboratoire de Recherche sur les Sciences de la Matière (LARSIM, CEA-Saclay) directed by Etienne Klein. He has worked as a theorist in the domain of particle physics (PhD in 1967, about sixty publications in international reviews) in the theory department (Service de Physique Théorique) and then in the experimental department (Service de Physique des Particules). At the level of the Master degree, he has taught theoretical physics at Orsay from 1979 to 1983, history of theoretical physics, also at Orsay from 2000 to 2004 and now, philosophy of physics at Paris. He is the author, or co-author of some books of reflection on physics directed to a wide audience: *La Matière-Espace-Temps*, with Michel Spiro in 1986; *L'Horizon des Particules*, with Jean Pierre Batou in 1989; *Les Constantes Universelles* in 1991, a book that has been translated in English (Mac Graw Hill). He is currently preparing a thesis in philosophy of science on the work of the Swiss philosopher Ferdinand Gonseth, under the direction of Dominique Lecourt.

Contact: LARSIM, CEA Saclay, F-91191 Gif-sur-Yvette cedex. gicotan@club-internet.fr

François Dagognet, former director of the Institut d'histoire et de philosophie des sciences et des techniques and emeritus professor at the Université de Paris I Panthéon-Sorbonne, has devoted himself to conceive materiality (as a materiologist not a materialist), of which he has studied different aspects. Among his recent books:

Philosophie du transfert (Encre marine, 2006); *Une épistémologie de l'espace concret* (reed., Vrin, 2006). In English translation: *Etienne-Jules Marey: A Passion for the Trace* (N.Y., Zone Books, 1992). Several books have been devoted to his work: G. Canguilhem et al., *Anatomie d'un épistémologue: François Dagognet* (Paris, Vrin, 1984); A. Métraux (ed.), *François Dagognet: Historian, Philosopher* (vol. 9 of *Science in Context*, March 1996); F. Dagognet and P. Vighetti, *Suivre son chemin: un itinéraire philosophique* (Genouilleux, La Passe du vent, 2006, initially issued as *Cheminement*, Paroles d'Aube, 1996); R. Damien (ed.), *François Dagognet, médecin, épistémologue, philosophe: une philosophie à l'oeuvre*, Le Plessis-Robinson, Synthélabo, 1997); G. Chazal and C. Salomon (ed.), *François Dagognet, médecin et philosophe* (Paris, l'Harmattan, 2006).

Contact: UFR de philosophie, Université de Paris I Panthéon-Sorbonne, 17 rue de la Sorbonne, F-75231 Paris Cedex 05.

Claude Debru studied philosophy at the École normale supérieure and biophysical chemistry at the Pierre et Marie Curie University in Paris. He worked under the supervision of Georges Canguilhem and François Dagognet in France, and of John Edsall and Jeffries Wyman at Harvard and Rome. His doctoral dissertation dealt with the history and philosophy of protein biochemistry (1982). He moved to the Department of Experimental Medicine at the Lyon Faculty of Medicine and worked on the current problems of sleep and dreaming neurophysiology with Michel Jouvet. He also interacted with the hematologists Jean Bernard and Marcel Bessis on leukemia classifications. He was the founding secretary of the European Association for the History of Medicine and Health and the founder of the European Society for the History of Science. He is a corresponding member of the French Academy of Science and a member of the Deutsche Akademie der Naturforscher Leopoldina. He is President of the French National Committee for History and Philosophy of Science. Main works: *L'Esprit des Protéines* (Paris, Hermann, 1983); *Neurophilosophie du rêve* (Paris, Hermann, 1989); *Philosophie de l'inconnu* (Paris PUF, 1998); *Le possible et les biotechnologies* (Paris, PUF 2003); *Georges Canguilhem : science et non-science* (Paris, Editions rue d'Ulm, 2004).

Contact: École normale supérieure, 45 rue d'Ulm, F-75005 Paris. claudio.debru@ens.fr

Anne Fagot-Largeault, PhD (Stanford 1971), MD (Paris 1978), Docteur ès lettres et sciences humaines (Paris 1986). Currently full professor of the Collège de France (philosophy of life science). Member of the French Academy of Sciences; of Academia Europea; of the International Academy of Philosophy of Science. Member of the French High Council for Science and Technology and of the College of Experts of the Biomedicine Agency. Has been (and still is) a member of Data and Safety Monitoring Committees for international clinical trials in aids, cancer, cellular therapy. Career: she worked as a full time professor in philosophy in high school, then as an assistant, assistant professor, full professor in philosophy of science at the universities of Paris-12, Paris-10, Paris-1; she did her inaugural lesson at the Collège de France in 2001. Her medical practice at the Henri Mondor hospital (Assistance publique de Paris, one full day a week, from 1978 on) was mostly as a

psychiatrist in the emergency room; her ongoing cooperation with the hospital psychiatry wards is now directed towards the methodological and ethical problems in psychiatric investigations. Her research, centered on the philosophy of life sciences, developed along three lines: (1) diagnostic reasoning, inductive logic, statistical and probabilistic methods, heuristic procedures; (2) investigation of causal links, validation of causal hypotheses, causal *vs.* teleological explanation; (3) ethics of medical practice and investigation, methodology of clinical & of epidemiological research, epistemology of bio-medical sciences, bio-medical anthropology. Her list of publications counts around 250 titles, including a dozen of books, among which: *Médecine et probabilités*, Paris: Didier-érudition, 1982; *L'homme bio-éthique. Pour une déontologie de la recherche sur le vivant*, Paris: Maloine, 1985; *Les causes de la mort. Histoire naturelle et facteurs de risque*, Paris: Vrin, 1989; with D. Andler et B. Saint-Sernin, *Philosophie des sciences*, 2 vols., Paris: Gallimard, 2002; and, with S. Rahman and J.M. Torres, *The Influence of Genetics on Contemporary Thinking*, Dordrecht: Springer, 2007. See list of publications, www.college-de-france.fr. Contact: Collège de France, 11 place Marcelin Berthelot, F-75005 Paris. anne.fagot-largeault@college-de-france.fr

Jean Gayon was born in 1949. Trained in philosophy and biology in Paris, he has been successively Professor at the University of Dijon, at Paris 7 University, and Paris 1 University, where he is member of the Institut d'histoire des sciences et des techniques (since 2001). A specialist of history and philosophy of biology, he has published 15 personal or collective books and around 200 articles or chapters in collective books, mainly on evolutionary biology, genetics and biometry, general philosophy of science, and social and ethical problems raised by the life sciences. In 1994, he was appointed senior member of Institut Universitaire de France. Other distinctions: Grammaticakis-Neuman Prize of Philosophy of Science (Paris Academy of Science, 2002); member of the German National Academy of Science (formerly "Leopoldina Academy, 2002); Corresponding member of the International Academy of History of Science (2006); member of the International Academy of Philosophy of Science (2007). Main publications: *Darwinism's Struggle for Survival* (Cambridge UP, 1998); *1900: Redécouverte des lois de Mendel* (ed. in coll. with F. Gros, M. Morange, M. Veuille, Ac. Sci. Paris, 2000); *L'épistémologie française 1830–1970* (ed. in coll. With M. Bitbol); *L'éternel retour de l'eugénisme* (ed. in coll. with D. Jacobi, Paris, PUF, 2006); *Conceptions de la science: Hier, aujourd'hui et demain. Hommage à Marjorie Grene* (ed. in coll. with R. M. Burian, Bruxelles, Ousia, 2007).

Contact: IHPST, 13 rue du Four, F-75006 Paris. gayon@noos.fr

Sandra Laugier is Professor of Philosophy (Epistemology) at Université de Picardie (Amiens) and former member of the Institut Universitaire de France. After studying at the Ecole Normale Supérieure (Paris) and at Harvard University (with B. Dreben, S. Cavell, H. Putnam), she has focused on philosophy of language and theory of meaning, first on philosophy of language and epistemology in Quine's work (her PhD was on Quine's indeterminacy thesis: *L'anthropologie logique de Quine*, Paris, Vrin, 1992), then, following J. Bouveresse, on a general "realistic" approach

of the relation of epistemology to ordinary language, as illustrated in Wittgenstein's and Austin's methods: *Du réel à l'ordinaire*, Vrin, 1999; *Wittgenstein, Métaphysique et jeu de langage*, PUF, 2001. Her present work extends this approach to the connection between ethics, knowledge, and language, as defining ordinary moral realism: *Le souci des autres* (with P. Paperman) Paris, EHESS, 2005; *Ethique, littérature, vie humaine*, PUF, 2006. Her aim has also been to introduce new varieties of American philosophy in France, by translating important philosophical works (Cavell, Quine, Diamond) and editing several collections (*Philosophie des sciences*, with P. Wagner, 2 vols, Vrin, 2004, *Philosophie du langage*, Vrin, 2008). Contact: Département de philosophie, Université Picardie Jules Verne, Chemin du Thil, F-80025 Amiens. sandra.laugier@u-picardie.fr

Philippe Mongin is a Directeur de recherche at CNRS (Centre National de la Recherche Scientifique) and a Professeur d'économie et de philosophie at the HEC School of Management. He studied philosophy and political science at École Normale Supérieure and Institut d'Études Politiques, and economics at King's College and the University of Cambridge. In 1978, he took a Ph.D. from École des Hautes Études en Sciences Sociales and entered CNRS in the economics division. He has since been working there, but also held teaching positions in France and abroad, the longer ones at Ecole Normale Supérieure (1981–1988), Université Catholique de Louvain (1991–1996), and the London School of Economics (1999–2001). He made briefer visits to Canadian and US universities.

His initial interest, as exemplified by his dissertation on Marx's *Grundrisse*, was the philosophy of economics as viewed from an intellectual history perspective. By the mid-1980s, after taking a mathematics degree, he shifted to the theories of rational choice and collective decision-making. One of his main research programmes consisted in enriching social choice theory with axiomatics of rational choice under uncertainty. Another programme involved applying philosophical logic to the epistemic foundations of the theories of games and decisions; it reflects a long-standing concern with the formal representations of the deliberative reasoning underlying human decisions.

In the 1980s, Philippe Mongin's philosophy of economics used neo-positivist constructs, as in the economic methodology of the time, but in the 1990s and 2000s, it deemphasized philosophy of science, and increasingly resorted to analytical philosophy to discuss topics that are not well explored vis-a-vis economics, like the role of value judgments and untestable principles in a scientific inquiry.

He has published a monograph, several collections, and a number of journal articles and chapters in collective volumes, both in French and English. Contact: GREG-HEC, 1 rue de la Libération, F-78350 Jouy-en-Josas. mongin@hec.fr

Daniel Parrochia, born in 1951, studied philosophy and linguistics at the University of Lyon, where he was a student of François Dagognet, himself a disciple of Gaston Bachelard. Between 1979 and 1990, he was a researcher at the Institut d'histoire et de philosophie des sciences et des techniques (CNRS) in Paris. Then he taught logic and philosophy of science as a professor in the universities of Toulouse

II and Montpellier III. Since 2003, he is a professor of logic and philosophy of science at the University of Lyon, and the director of the Centre d'Études des Formes (Form Studies Center) in the Institute of philosophy of Lyon. Areas of research: Philosophy of technology. mathematical aspects of forms, structures, networks in society and culture. General theory of classifications. Selected publications: *Philosophie des Réseaux*, Paris, Presses Universitaires de France, 1993; "A Historical Note on Artificial Life", *Acta Biomathematica*, 43, 1995, 177–183; "Algorithms and the Limits of Complexity", *Science in Context* 9, 1, 1996, pp. 39–56; *La conception technologique*, Paris, Hermès, 1998; *L'homme volant, philosophie de l'aéronautique et des techniques de navigation*, Seyssel, Champ Vallon, 2003.

Contact: Faculté de philosophie, Université Jean Moulin Lyon 3, 18 rue Chevreul, F-69007 Lyon. daniel.parrochia@univ-lyon3.fr

Joëlle Proust studied philosophy and psychology at the University of Provence. A researcher at Centre national de la recherche scientifique since 1976, she first studied the history and the philosophy of logic. Her first book, *Questions of form* (Gallimard, 1986, Minnesota Press, 1989) deals with the status of analytic propositions from Kant to Carnap. It received the bronze medal of the CNRS. From then on, her main interest is in the analytic philosophy of mind: she published articles and books on intentionality and animal cognition (*Comment l'esprit vient aux bêtes*, Paris, 1997; *Les animaux pensent-ils?* 2003), on the kind of awareness involved in agency and in personal identity and in its perturbations in schizophrenia and autism. Her last book, entitled *La nature et la volonté*, was published in 2005. It defends a volitionist conception of the intentional contents involved in action. She presently directs a European Science Foundation Collaborative Research Project on metacognition, i.e. the capacity to predict or evaluate the cognitive adequacy of one's own mental states, a capacity highly relevant to understanding the evolutionary basis of epistemology and self-consciousness.

Contact: Institut Jean Nicod, CNRS UMR 8129, 29 rue d'Ulm, F-75005 Paris. jproust@ehess.fr

Hervé Zwirn is a graduate of the École polytechnique, holds a telecommunications engineering degree from Télécom Paris, and a PhD in particles physics from Paris VI University. He is currently Associated Director of Research at the CNRS and at the Ecole Normale Supérieure de Cachan, and associated researcher at the Institut d'histoire et philosophie des sciences et des techniques (Paris). Hervé Zwirn started his research in Quantum Physics before moving to problems raised by the interpretation of quantum mechanics and then to general epistemology. He has been particularly interested in the problem of realism and in the impact of modern physics on the vision of the world. He has published a book of philosophy of science on this subject: *Les limites de la connaissance*, Odile Jacob, 2000. He has also worked in the domain of modelling human reasoning, inductive logic and belief revision. He has proposed a new modelling of preferences in decision theory based on the quantum formalism. His current main domain of interest is the Sciences of Complexity for which he has recently published a book: *Les Systèmes Complexes*, Odile Jacob, 2006. Among his other publications are: "Abductive Logic in a Belief Revision

Framework”, with B. Walliser and D. Zwirn, *Journal of Logic, language and Information*, 14, 1, 2005; “Can we Consider Quantum Mechanics to be a Description of Reality ?”, in *Rethinking Scientific Change and Theory Comparison. Stabilities, Ruptures, Incommensurabilities*, L. Soler, P. Hoyningen, H. Sankey (eds), Springer (to appear). “Type Indeterminacy: A Model of the KT (Kahneman-Tversky)-Man”, with A. Lambert et S. Zamir, *Journal of Economic Theory* (to appear).

Contact: Centre de Mathématiques et de leurs applications – UMR 8536, École Normale Supérieure de Cachan, 61, avenue du Président Wilson, F-94235 Cachan cedex. herve.zwirn@m4x.org

Name Index

A

Abdellaoui, M., 350
Adler, F., 102
Agazzi, E., 162, 163
Alberch, P., 206
Alembert, J. Le Rond d', 2, 18, 30, 44, 47, 52, 80, 182
Allais, M., 306, 324–342, 345–349, 352–356
Althusser, L., 213, 215
Amit, D., 296
Ampère, A.M., 3, 5
Amundson, R., 206, 250, 251
Anderson, J., 296, 299
Andler, D., 14–17, 46, 255, 293–296, 359, 362
Arbib, M., 239, 252
Arendt, H., 61
Ariew, R., 111, 350, 354
Aristotle, 10, 58, 75, 84, 173–176, 180, 183–185, 231, 261, 267
Aserinsky, E., 216
Aspect, A., 131, 158
Atran, S., 45
Austin, J.L., 363

B

Babbage, C., 294
Bachelard, G., 1, 2, 5–8, 10, 11, 13, 14, 19–22, 26, 35, 36, 44–46, 56, 63, 64, 67, 68, 74, 76–79, 81, 83, 85, 87, 88, 167, 168, 173, 174, 178, 179, 182–184, 206, 214, 304, 363
Bachelard, S., 213, 214, 223, 224
Bacon, F., 29, 30, 44, 46, 53
Bailhache, P., 12
Bain, A., 272, 273, 279, 294–296
Balfour, A., 172
Barbara, J.G., 45, 46, 112
Barberousse, A., 139
Barkow, J., 295, 297

Barnes, B., 45
Barsalou, L.W., 250, 251
Barthes, R., 59, 68
Baton, J.P., 360
Baudrillard, J., 16, 67, 68
Bayle, F., 67, 68, 70
Beatty, J., 207
Beaune, J.C., 54, 64, 65, 67, 68, 200
Becquerel, H., 120
Bell, J., 22, 131, 158
Benjamin, W., 49, 60
Benner, S.A., 46
Bennett, M.R., 296, 297
Benoist, J., 20
Bensaude-Vincent, B., 14, 15, 18, 165, 182–184, 359, 360
Bergson, H., 4, 11, 17, 19–21, 37, 45, 46, 53, 54, 56, 67, 68, 77
Berkeley, G., 48, 49, 89, 144, 153, 186, 298, 359
Bermudez, J.L., 295
Bernard, C., 3, 18, 36, 54, 70, 163, 185, 194–200, 214, 217, 218, 304, 361
Bernard, J., 215, 220, 221, 223, 224
Bernoulli, D., 327
Berson, J.A., 182, 184
Berthelot, M., 25, 40, 46, 168, 182, 362
Berthollet, C., 170, 185
Berthoz, A., 293, 297
Bessis, M., 215, 220, 223, 224, 361
Besso, M., 102, 138
Biot, J.B., 314
Bird, A., 143, 249, 251, 281
Bitbol, M., 20, 21, 108, 110, 163, 362
Blanché, R., 12
Bloch, M., 57
Block, N., 143, 275, 295, 297, 316
Bloor, D., 45
Boerhaave, H., 179

- Bohm, D., 131
 Bohr, N., 103, 158, 177, 183, 184
 Boltzmann, L., 121–123, 137
 Bolzano, B., 2, 20
 Bonatti, L., 295, 297
 Bonnet, C., 110
 Boole, G., 294
 Born, M., 8, 9, 11, 63, 104, 128, 286, 341,
 360, 362, 363
 Bose, S., 130, 139
 Boudot, M., 12
 Bourdeau, M., 20
 Bourdieu, P., 16
 Bourg, D., 54, 67, 68, 70
 Boutroux, É., 3, 4, 10
 Bouveresse, J., 8, 9, 20, 21, 362
 Bouyssou, D., 352, 354
 Boyd, R.T., 146, 156, 162, 163, 251
 Boyer, A., 110, 350, 355
 Boyle, R., 27, 171, 174, 185, 186
 Braddon-Mitchell, D., 293, 297
 Braunstein, J.F., 19, 21
 Brenner, A., 1, 13, 15, 16, 18, 20, 21, 73, 88,
 108–111, 350, 351, 355, 360
 Brentano, F., 261
 Breton, P., 54, 60, 67, 68
 Bridgman, P.W., 101
 Brittan, G., 20, 21
 Broca, P.P., 239
 Brodmann, K., 239
 Broglie, L. de, 18, 128, 129, 139
 Brooke, A., 293
 Brunner, J., 246
 Brunschvicg, L., 2, 10, 35, 39, 45, 46, 78, 184
 Buffon, G., 80, 81, 88
 Buller, D.J., 250, 251
 Burian, R., 186, 204, 206, 207, 362
 Burkitt, G.L., 223
 Busen, R.W.,
 Buser, P., 215
 Butler, S., 58
 Buytendijk, F.J.J., 38
- C**
- Caligaris-Cappio, F., 222, 224
 Camerer, C., 325, 335, 350, 353–355
 Campbell, J., 206
 Canguilhem, G., 2, 5, 7, 8, 10, 11, 19–22, 34,
 35, 45–47, 54, 55, 67, 68, 184, 185,
 200, 203, 206, 212, 214, 215, 223,
 224, 304, 361
 Carnap, R., 9, 33, 83, 92, 93, 95, 97, 99, 109,
 111, 364
- Carnot, S., 53, 68, 121
 Carroll, L., 222
 Carruthers, P., 295, 297, 300
 Carter, D., 249, 251
 Cassirer, E., 178, 183, 185, 261
 Cavailès, J., 8, 12, 20
 Cavell, S., 362, 363
 Cayley, G., 66
 Chaisson, E.J., 38, 46, 47
 Chaitin, G., 153, 162, 163
 Chalmers, D., 275, 293, 295, 297
 Changeux, J.P., 242, 243, 250, 251, 293, 297
 Charpentier-Morizé, M., 183, 185
 Chazal, G., 54, 65, 67, 68, 361
 Chew, S.H., 66, 353, 355
 Chomsky, N., 252, 257, 272, 278, 291,
 295–297, 299
 Christensen, W.D., 250, 251
 Clark, A., 34, 296, 297
 Clausius, R., 121, 139
 Clavelin, M., 12
 Clericuzio, A., 183, 185
 Cohen, I.B., 36, 89
 Cohen, M., 350, 353, 355
 Cohen, R.S., 18, 21, 22, 89, 224, 229, 299
 Cohen-Tannoudji, G., 13, 17, 115, 139, 360
 Compton, A.H., 129
 Comte, A., 2–5, 9, 10, 15, 16, 18–21, 30, 34,
 44, 47, 52, 53, 64, 67, 68, 76, 78,
 82, 83, 88, 101, 165, 177, 201, 202,
 206, 212, 214
 Condillac, E. de, 175, 176, 183, 185
 Condorcet, N. de, 52, 68
 Copernicus, 74, 75, 81, 86, 150
 Coriolis, G.G., 53, 68
 Cosmides, L., 237, 250, 251, 295, 297
 Couffignal, L., 54, 68
 Cournot, A.A., 3, 30, 31, 34, 36, 44, 45, 47,
 49, 78
 Couturat, L., 11
 Cowie, F., 295, 297
 Cozic, M., 295, 297
 Craig, E., 293, 297
 Crombie, A.C., 45
 Curie, M., 121
 Curie, P., 121
- D**
- Dagognet, F., 8, 11, 14, 15, 17, 19–21, 54,
 63–68, 189, 360, 361, 363
 Dalton, J., 176
 Damien, R., 21, 361
 Darwin, C., 22, 30

Daston, L., 36
 Daumas, M., 54, 55, 57, 67, 68
 Davidson, D., 33, 297
 Davies, M., 293–295, 297
 Dawkin, R., 250
 Debray, R., 54, 67, 68, 70
 Debru, C., 14, 15, 17, 41, 46, 47, 213,
 224, 361
 Dehaene, S., 242, 243, 250, 251
 Dehaene-Lambertz, G., 250, 251
 Dekeuwer, C., 46, 47
 Del Re, G., 182, 185
 Delacampagne, C., 108
 Deleuze, G., 16
 Democritus, 173
 Dennett, D., 272, 274, 293, 295, 297
 Derrida, J., 45, 213, 223, 224
 Descartes, R., 8, 10, 27, 39, 44, 47, 51, 52,
 66, 68, 75, 80, 81, 87, 88, 173, 261,
 266, 292
 Destouche-Février, P., 20
 Dewan, E., 218, 223, 224
 Dewey, J., 101
 Diamond, C., 363
 Diderot, D., 2, 18, 44, 47, 52, 80, 163, 168,
 173, 182
 Diecidue, E., 352, 355
 Dijksterhuis, E.J., 35
 Dilthey, W., 27, 31, 44, 47
 Dingler, H., 212
 Dirac, P., 128, 166, 186
 Dobbs, B.J., 183, 186
 Dreben, B., 362
 Dretske, F., 271, 294, 297
 Drexler, E., 181, 184, 185
 Dreyfus, H., 273, 291, 293, 295, 296, 298
 Dubucs, J., 12, 296
 Duhem, P., 2, 4–6, 10, 11, 13, 14, 20, 21, 34,
 35, 45, 47, 76, 82, 88, 89, 91–95,
 97–102, 104, 105, 108–111, 166,
 174, 176, 182–185, 303–314, 316,
 331, 333, 341, 342, 345–347, 350,
 351, 354–357, 360
 Dummett, M., 163
 Duncan, J., 250, 251, 356
 Dupoux, E., 295, 298
 Dupuy, J.P., 67, 68

E
 Ebinger, M., 64
 Eddington, A.S., 101, 294
 Edelman, G.M., 242, 251
 Edsall, J., 215, 361

Egrè, P., 45, 47
 Einstein, A., 76, 81, 101–103, 109–111,
 115–117, 122–132, 134, 135,
 138–140, 150, 151, 155, 158, 214
 Elliot, J., 110
 Ellsberg, D., 319, 321, 322, 352, 355
 Ellul, J., 56, 57, 63, 68, 69
 Elman, J., 295, 298
 Empedocles, 173
 Engels, F., 59
 Epicurus, 173
 Erneling, C., 293, 298
 Espagnat, B. d', 18, 144, 145, 148, 162, 163
 Ettigoffer, D., 68
 Euclid, 33
 Euler, L., 118, 119

F
 Fagot-Largeault, A., 13, 17, 18, 25, 46,
 361, 362
 Faraday, M., 123
 Farben, I.G., 197
 Febvre, L., 57
 Feigl, H., 261
 Feselstein, J., 206
 Feyerabend, P., 21, 88, 91, 98–100, 104, 105,
 156, 162, 163, 360
 Feynman, R., 133–135, 139
 Fields, J.C., 6, 16, 56, 65, 119, 123, 126,
 129–136, 140, 160, 185,
 255, 256
 Fine, A., 103, 110, 111, 136, 177, 241, 287
 Finetti, B. de, 324, 348, 352
 Finkielkraut, A., 68
 Fishburn, P.C., 332, 351–353, 355
 Fleck, L., 35, 36, 43, 45–47
 Fleming, A., 197, 198
 Fodor, J., 232, 251, 269, 275, 279, 280, 294,
 295, 298, 350, 355
 Foucault, J.B.L., 308, 310, 311, 314,
 343, 351
 Foucault, M., 2, 8, 10, 16, 19–21, 36, 45, 47,
 85, 108
 Fourier, C., 58
 Fraenkel, A.A., 158
 Frank, P., 22, 49, 54, 68, 70, 92, 110
 Frege, G., 267, 268
 Fresnel, A., 311, 312, 314
 Freud, S., 37, 45, 47
 Friedman, D., 306, 318, 322–324, 335, 343,
 352–355
 Friedmann, G., 54, 69
 Frith, C., 295

Fromherz, S., 223, 224
Fruteau de Lacos, F., 110

G

Gabor, D., 57, 63
Galen, 191, 192
Galileo, 10, 55, 102, 117, 171
Gall, F., 279
Gärdenfors, P., 162, 163
Gardies, J.L., 12
Gateau, V., 46, 47
Gates, B., 60
Gavroglu, K., 13, 18, 183, 186
Gayon, J., 1, 14, 15, 17, 19–21, 65, 108, 110, 115, 201, 362
Geber, pseudo, 183, 185
Gell-Mann, M., 139, 140
Gelman, S.A., 295, 298, 300
Gerhardt, C., 169
Ghiselin, M., 207
Gibbs, G.W., 122, 123, 137, 139, 140
Gigerenze, G., 295, 298
Gilbert, W., 11, 22, 41, 46, 48, 54, 55, 57, 64, 69, 70, 261
Gilboa, I., 353, 355
Gille, B., 56, 57, 67, 69
Gillies, D., 350, 351, 355
Glashow, L., 149
Gley, E., 212
Glymour, C.C., 34
Gödel, K., 65, 88, 145, 158, 268, 294, 299
Goffette, J., 46, 47
Goffi, J.Y., 54
Goldstein, K., 202
Gonseth, F., 138–140, 360
Goodman, B., 206
Gould, S.J., 230
Granger, G.G., 8–10, 22, 353, 355
Gras, A., 54, 67, 69
Grene, M., 19, 21, 203, 204, 207, 362
Griesemer, J.R., 206
Griffiths, P.E., 235, 237–240, 250–253
Gros, F., 362
Grosholz, E.R., 182, 185
Grünbaum, A., 33, 89, 309, 351, 355, 356
Guala, F., 352, 353, 355, 356
Gueroult, M., 8
Guillerme, A., 54, 69
Gutting, G., 19–22
Guyton de Morveau, L.B., 183, 185
Gzil, F., 46, 47

H

Haak, S.,
Habermas, J., 59, 69

Hacker, P.M.S., 296, 297
Hacking, I., 21, 36, 45, 47, 48, 85–89, 98, 99, 105, 109, 111, 179, 180, 183, 185, 360
Hagen, O., 334, 335, 353, 354, 356
Hahnemann, S., 193, 198
Haller, A. von, 195
Hallyn, F., 80, 88, 89
Hamelin, O., 104, 112
Hamilton, W., 104, 112, 118, 119, 272, 294
Hammerstein, P., 250, 252
Hammond, P.J., 352, 356
Hannequin, A., 5
Hanson, N.R., 162, 163
Hansson, S.O., 33
Harding, S.G., 350, 351, 355, 356
Harman, P.M., 296, 298
Hartfield, G.,
Hartle, J.B., 139, 140
Harvey, W., 196
Haudricourt, A.G., 57
Haugeland, J., 89, 111, 293, 298
Hauser, M., 234, 252
Hausman, D., 353, 356
Heaviside, O., 123
Hegel, G.W.F., 31, 52, 57, 60, 104, 112, 144, 206
Heidegger, M., 26, 44, 48, 54, 59, 62, 63, 67, 69, 213
Heidelberger, M., 20, 21
Heil, J., 293, 298
Heisenberg, W.K., 103, 116, 128, 130, 131
Helmholtz, H. von, 75, 272
Helmont, J.B. von, 167, 182
Hempel, C.G., 83, 84, 88, 89, 92
Henderson, L., 214
Hendricks, V.F., 45, 48
Henry, M., 53, 57, 69
Herbrand, J., 12
Herken, R., 294, 298
Hermitte, M.A., 68
Herschel, J.F.W., 30, 44, 48
Herstein, I., 318
Hertz, H., 123
Hey, J.D., 353, 356
Hilbert, D., 129, 133
Hill, B., 47, 350, 360
Hintikka, J., 33
Hippocrates, 191, 192, 198
Hirschfeld, L., 295, 298, 300
Hobbes, T., 294
Hoffmann, R., 167–170, 179, 182, 183, 185
Hofstadter, D., 272, 294, 295, 298
Holmes, F.L., 167, 182, 185
Holton, G., 102, 110, 111

Hooker, C.A., 250, 251
 Horgan, T., 294, 296, 298, 371
 Hottois, G., 41, 46, 48, 54, 57, 69
 Howson, C., 351
 Hoyningen, P., 163, 365
 Hull, D., 186, 203, 204, 206–208, 212
 Hume, D., 64, 80, 87, 89, 144, 146, 147, 155
 Husserl, E., 11, 20, 28, 29, 36, 44, 48,
 213–215, 223, 224, 257, 261, 298
 Huxley, A., 58
 Huygens, C., 36, 37, 45, 48, 66

I

Illich, I., 63
 Imbert, M., 293, 298

J

Jackendoff, R., 234, 252
 Jackson, F., 293, 297, 298
 Jacob, F., 18, 48, 68, 69, 163, 224, 297, 298,
 300, 355, 356, 364
 Jacobi, C.G., 118
 Jacobi, D., 362
 Jacomy, B., 54, 68, 69
 Jaenisch, R., 249, 251
 Jaffray, J.Y., 353, 356
 Jallais, S., 352, 356
 James, W., 4, 19, 77, 111, 140, 185, 212,
 252, 296
 Janicaud, D., 54, 67–69
 Janik, A., 19, 22
 Jeannerod, M., 293, 298
 Jenner, E., 193
 Johnson-Laird, P.N., 281, 295, 298
 Johnston, J., 212
 Jonas, H., 61, 203
 Jordan, P., 128
 Jordan, S., 249
 Joule, J.P., 53
 Jouvét, M., 215–219, 223, 224, 361

K

Kahneman, D., 299, 331, 332, 335, 336, 342,
 345, 346, 353, 356, 365
 Kalinowski, G., 12
 Kant, I., 3, 4, 8, 10, 28, 31, 35, 44, 48, 142,
 145, 146, 168, 182, 261, 293, 294,
 297, 299, 364
 Kapp, E., 52
 Karmakar, U., 331, 353
 Karmiloff-Smith, A., 244, 248, 250, 252, 298
 Karni, 353, 355
 Katz, L., 250, 252
 Kauffman, S.A., 206
 Kekulé, F. A. von Stradonitz, 177

Kepler, J., 45, 48, 74, 89, 150
 Kitcher, P., 293, 299
 Kitcher, Ph., 291, 293, 296, 299
 Klein, E., 360
 Klein, U., 183–185
 Kleitman, N., 216
 Kneser, A., 101
 Knoblich, G., 249
 Koechlin, E., 250, 251
 Kolmogorov, A., 153
 Kotobi, H.K., 46, 48
 Koyré, A., 6, 10, 34, 35, 58, 59, 91, 100,
 105–108
 Kreisel, G., 20
 Krivine, J.L., 20
 Krüger, L., 35, 36
 Kuhn, T., 9, 13, 16, 21, 32, 35, 36, 43, 45, 48,
 74–77, 81–84, 86–89, 91, 98–100,
 104–108, 110, 111, 162, 163, 175,
 181, 183–185, 276, 290, 341, 347,
 360

L

Lacan, J., 213
 Lacoste Lareymondie, M. de, 46, 48
 Ladrrière, J., 67, 69
 Laffitte, J., 54, 69
 Lagrange, J.L., 88, 89, 118, 119
 Lajtha, L., 221, 222, 224
 Lakatos, I., 98, 99, 109–111, 305, 316, 341,
 345–347, 350, 351, 354–356
 Lalande, A., 182
 Lamarck, J.B. de, 27, 30, 44, 48
 Lambert, A., 365
 Lande, R., 206
 Langer, S., 261
 Laplace, P.S. de, 157
 Lardreau-Cotelle, E., 45, 48
 Largeault, J., 13, 17, 18, 25, 37, 45, 46, 48,
 361, 362
 Larsson, S., 329–331, 346, 349, 350, 352, 356
 Lashley, K., 272
 Latil, P. de, 54
 Latour, B., 11, 20, 22, 45, 63, 67–69, 183
 Laudan, L., 22, 49, 88, 89, 351, 356
 Lauder, G.V., 206, 250, 251
 Laufer, R., 68
 Laugier, S., 13, 16, 18, 91, 109, 111, 163, 351,
 356, 362, 363
 Lautman, A., 12, 37
 Lavoisier, A.L., 27, 36, 40, 105, 175, 176, 183,
 185, 359
 Le Chatelier, H.L., 53, 67, 69
 Le Roy, É., 4, 5, 19
 Lebeau, A., 69

- Lechopier, N., 46, 48
 Lecourt, D., 11, 19, 20, 22, 69, 200, 356, 360
 Legendre, G., 295, 296, 300
 Lehn, J.M., 183, 185
 Lehoucq, R., 139, 140
 Leibniz, G.W., 118, 267, 272, 294
 Lemaître, G., 138–140
 Lemery, N., 169
 Lepore, E., 355
 Leroi-Gourhan, A., 54–57, 67, 69
 Lesseps, F. de, 53
 Levi Strauss, C., 54, 69, 213
 Levine, J., 295, 299
 Lévy, M., 182, 185
 Lévy, P., 68, 69
 Lévy-Bruhl, L., 106, 107, 110–112
 Lewontin, R., 206, 230
 Liegener, C., 182, 185
 Linqvist, S., 180, 184
 Livingston, P., 293, 299
 Locke, J., 149, 266
 Loewi, O., 196
 Loomes, G., 353, 356
 Lopez, S., 295, 300
 Lorentz, H.A., 101, 102, 124, 125
 Lorne, M.C., 227, 249
 Lovitt, W., 67, 69
 Lubbock, J., 157
 Lucas, J.R., 294, 299
 Luce, R.D., 318, 319, 321, 351, 356
 Ludwig, K., 44, 49, 293, 299
 Luppi, P.H., 217
- M**
- Mac Luhan, M., 52, 69
 MacCrimmon, K., 329–331, 346, 349, 350, 352, 356
 Macdonald, C., 296
 MacDonald, G., 296
 Mach, E., 19, 35, 45, 48, 76, 78, 82, 94, 95, 101–103, 111, 122, 214, 223, 224
 Machina, M.J., 332, 336–338, 342, 344, 350, 352–354, 356
 Magendie, F., 194
 Maitte, B., 54, 69
 Marcel, A.J., 69, 215, 220, 223, 224, 295, 297, 299, 361
 Marcel, G., 39
 Marcus, R.B., 34
 Marcuse, H., 64
 Marey, E.J., 361
 Markowitz, H., 352, 356
 Marr, D., 233, 240–242
 Marraffa, M., 293
 Marschak, J., 305, 306, 318, 320–324, 328, 332, 348, 353, 356
 Marshall, A., 69, 323, 352, 356
 Martin, R., 12, 26, 44, 48, 69, 213, 294, 297, 360
 Marx, K., 59, 63, 64, 363
 Massé, P., 352, 353
 Matthen, M., 238
 Maupertuis, P.L., 118
 Mauss, M., 57, 69
 Maxwell, J.C., 75, 78, 89, 120–124, 139, 140
 Maynard Smith, J., 206
 Mayr, E., 35, 45, 48, 207, 208, 228, 250, 252
 Mayrian, 288
 McClellan, E., 352, 356
 McCord, M., 331, 356
 McCrimmon, K.R., 329–331, 346, 349, 350, 352, 356
 McCulloch, W., 284, 293, 296, 299
 McIntyre, A., 43, 46, 48, 184, 186
 McLaughlin, P., 212
 Mendel, G., 362
 Mendeleev, D.I., 173, 178
 Merleau-Ponty, M., 257, 261
 Methot, P.O., 211
 Métraux, A., 361
 Metzger, H., 14, 106, 110, 112, 166, 184
 Meyerson, É., 2, 5–7, 10, 13, 14, 19, 35, 45, 48, 59, 91–93, 95–112, 166, 172, 176–178, 182–185
 Michelet, J., 58
 Michelson, A.A., 124
 Mignot, É., 218, 223, 224
 Milhaud, G., 4–6, 10, 19, 22, 82
 Mill, J.S., 4, 19, 22, 31, 87, 89
 Millikan, R., 252, 271, 294, 299
 Milnor, J., 318
 Minkowski, H., 124
 Mitcham, C., 67, 69
 Mithridates, 193
 Moles, A., 54, 68, 69
 Mongin, P., 14, 16, 303, 306, 351–354, 356, 357, 363
 Monod, J., 18, 38, 46, 48, 257
 Morange, M., 362
 Morgenstern, O., 305, 316–318, 322–324, 335, 349, 352, 355, 357
 Morlat, G., 352, 353
 Morley, E.W., 124
 Morton, P., 293, 299
 Mosteller, F., 323, 346, 347, 357
 Müller-Lyer, F., 232
 Mulliken, R., 178
 Mumford, L., 63, 64, 69, 70

Munier, B., 350, 354, 357
 Musgrave, A., 110, 111, 356
 Musso, P., 54, 60, 67, 70

N

Nagel, T., 295, 299
 Nani, M., 293
 Nash, J., 31
 Nativelle, C., 193
 Needham, P., 175, 183, 185
 Neufville, R. de, 356
 Neumann, J. von, 284, 285, 305, 308, 310,
 316–318, 322, 323
 Neurath, O., 5, 19, 22, 88, 94, 296, 299, 360
 Newcomen, T., 64
 Newell, A., 291, 296, 299
 Newman, W.R., 182, 183, 185, 186
 Newton, I., 10, 27, 29, 44, 48, 80, 81, 87, 89,
 103, 116–118, 120, 123, 126, 127,
 149–151, 157, 290
 Newton-Smith, W., 162, 163
 Nicod, J., 12, 20, 227, 364
 Nietzsche, F., 199, 200
 Nisbett, R., 295, 299
 Noether, E., 120
 Noguee, P., 323, 346, 347, 357
 Nordman, A., 183
 Nozick, R., 45

O

Ockham, W. of, 100
 Ostwald, W., 122, 170
 Oyama, S., 249, 252

P

Paillard, J., 216
 Paley, T., 30, 44, 48
 Panksepp, J., 238, 250, 252
 Pap, A., 32, 45, 48
 Paperman, P., 363
 Papineau, D., 271, 294, 299
 Pappus, 80
 Parenthesis, J., 224
 Pariente, J.C., 12
 Parot, F., 293, 299
 Parrochia, D., 13, 17, 51, 54, 67, 70, 363, 364
 Pascal, B., 51, 67
 Pasteur, L., 8, 21, 291
 Paty, M., 138, 140
 Pearson, K., 30, 44, 49
 Peirce, C.S., 268
 Pericles, 55
 Perrin, J., 68, 123, 185

Peschard, I., 110
 Petitot, J., 293
 Petzoldt, J., 110
 Piaget, J., 243, 257, 279, 299
 Piatelli-Palmarini, M., 293, 295
 Pickering, A., 45
 Picon, A., 68
 Picot, C., 80
 Pinker, S., 234, 252
 Planck, M., 116, 127–130, 136, 137
 Plato, 28, 62, 181, 192, 200, 261, 293
 Podolsky, B., 130, 139
 Poincaré, H., 2, 4–6, 11, 76, 82, 124, 360
 Poncelet, J.V., 52, 67, 70
 Popper, K., 9, 75, 83, 91, 99, 107, 149, 150,
 155, 157, 305, 307, 311–316, 331,
 345, 346, 351, 354, 355, 357
 Pradier, P.C., 352, 356
 Premack, D., 257
 Proudhon, P.J., 58
 Proust, J., 14, 16, 20, 22, 227, 235, 249–252,
 364
 Ptolemy, 74, 75
 Pullum, G., 295
 Pulte, H., 19, 22
 Putnam, H., 91, 98, 103, 143, 146, 152, 157,
 162, 163, 265, 266, 279, 291, 294,
 296, 299, 362
 Pylyshyn, Z.W., 269, 294

Q

Quinton, T., 34
 Quartz, S.R., 250, 252, 295, 299
 Quéau, P., 68
 Quiggin, J., 332, 338–340, 353, 357
 Quine, W.V., 4, 9, 13, 14, 91–101, 103, 104,
 106–109, 111, 112, 149, 151, 303,
 304, 308, 309, 311, 350, 351,
 355–357, 362, 363

R

Rabinow, P., 19, 22
 Rahman, R., 362
 Raiffa, H., 318, 319, 321, 351, 356
 Ramunni, G., 70
 Raup, D., 206
 Rees, G., 295
 Reichenbach, H., 6, 9, 33, 45, 49, 92
 Rémond, A., 216
 Renan, E., 9, 20, 22
 Renn, J., 13, 18
 Renouvier, C., 4, 19, 22
 Rescher, N., 162, 163

- Reuleaux, F., 53
 Rey, A., 2, 4–6, 21, 78
 Rey, G., 293
 Rheinberger, H.J., 19, 22
 Richelle, M., 293, 299
 Ricœur, P., 16, 84–86, 88, 89, 213, 223
 Rips, L.J., 295, 299
 Ritter, K., 22, 52
 Rizzolatti, G., 239, 252
 Roentgen, W.K., 120
 Rosen, N., 130, 139, 140
 Rosenberg, A., 212
 Rosenfeld, E., 184, 185, 296
 Rosenthal, D.M., 293, 299
 Rougier, L., 107, 111, 112
 Rouilhan, P. de, 12
 Roy, A., 4, 5, 19, 82, 239, 252, 299
 Rubin, H., 352, 357
 Rumelhart, D., 295, 299
 Ruse, M., 204, 207, 212
 Russell, B., 9, 19, 28, 29, 33, 44, 45, 49, 145,
 261, 325
 Russo, F., 54, 70
 Ruyer, R., 67, 70
 Ryle, G., 261
- S**
 Safra, 353, 355
 Saint-Sernin, B., 43, 45, 46, 49, 362
 Saint-Simon, C.H. de, 52, 60, 64, 70
 Salam, A., 149
 Sallanon, M., 223, 224
 Salomon, C., 361
 Salomon, J.J., 54, 70
 Samuelson, P.A., 306, 318, 322, 324, 348,
 351–354, 357
 Sandu, G., 34
 Sankey, H., 163, 351, 365
 Sarton, G., 7
 Sartre, J.P., 39
 Sastre, J.P., 223, 224
 Savage, L.J., 305, 306, 317, 318, 322–325,
 327–329, 334, 335, 340, 343, 348,
 349, 352, 354, 355, 357
 Savery, T., 64
 Sawyer, K., 351
 Schaffner, K., 212
 Schelling, F.W.J. von, 31
 Scheps, R., 68
 Schlick, M., 20, 101, 103, 109, 110, 112
 Schmeidler, D., 340, 353, 357
 Schmidt, U., 352, 355
- Schneider, S., 295, 300
 Scholz, B., 295
 Schopenhauer, A., 31
 Schrödinger, E., 17, 37, 103, 128, 129, 131,
 134
 Schuhl, P.M., 54, 58, 59, 67, 70
 Schwartz, E., 20, 22
 Searle, J., 291, 295, 296, 299
 Segal, U., 354, 357
 Sejnowski, T., 250, 252, 295, 299
 Sellars, W., 143, 162, 163, 261
 Selten, R., 295, 298
 Séris, J.P., 54–57, 60, 63, 66, 67, 70
 Serres, M., 2, 19, 21, 22, 54, 57, 70, 359
 Sfez, L., 54, 60, 67, 70
 Shank, R., 286
 Shatz, C., 250, 252
 Shockley, K., 250, 252
 Sicard, M., 68
 Sigaut, F., 70
 Simon, H., 52, 64, 70, 88, 182, 243, 248, 252,
 282, 286, 291, 295, 296, 299, 300,
 349, 359
 Simon, J., 182, 359
 Simon, K., 252
 Simondon, G., 11, 20, 22, 54, 64, 67, 70
 Sismour, A.M., 46
 Slovic, P., 299, 352, 357
 Smith, D.W., 293, 298
 Smith, L., 250, 253
 Smith, M., 297, 298
 Smith, V.E., 212
 Smolensky, P., 272, 283, 285, 294–296, 300
 Smolin, L., 139, 140
 Snyder, C., 295, 300
 Sober, E., 206
 Soler, L., 162, 163, 365
 Solomonoff, R., 153
 Sommerfeld, A., 101, 102
 Spinoza, B., 39
 Spiro, M., 360
 Stadler, F., 20, 21
 Stainton, R., 295, 300
 Starmer, C., 350, 357
 Starobinski, J., 85, 86, 88, 89
 Stengers, I., 14, 170, 172, 182, 184, 186, 359
 Sterelny, K., 235, 237, 248–251, 253
 Stich, S., 272, 282, 295, 297, 299, 300
 Stiegler, B., 54, 68, 70
 Sugden, R., 350, 353, 356, 357
 Sunder, S., 353, 355
 Suppes, P., 34, 39, 45, 46, 49

Susskind, L., 139, 140
 Symons, J., 45, 48
 Szubka, T., 293, 300

T

Taine, H., 53
 Tallon, J.M., 350, 353, 355
 Tannery, P., 3, 47, 88
 Tarski, A., 92
 Thelen, E., 229, 250, 253
 Thom, R., 18, 38, 46, 49, 65, 257, 293, 300
 Thomasson, A.L., 293, 300
 Thomson, J.J., 120
 Tienson, J., 294, 296, 298
 Tinland, F., 54, 68, 70
 Tommasi, L., 251
 Tooby, J., 237, 250, 251, 297
 Torres, J.M., 362
 Toulmin, S., 19, 22, 27, 44, 49
 Turing, A., 158, 265, 266, 268, 269, 277, 278,
 283–285, 294–296, 298, 300
 Tversky, A., 299, 331, 332, 335, 336, 342,
 345, 346, 352, 353, 356, 357, 365

U

Uexküll, J.J. von, 38
 Urbach, P., 46, 351
 Uzan, J.P., 139, 140

V

Valéry, P., 27, 44, 49
 Van Fraassen, B., 156, 162, 163
 Vansnick, J.C., 352, 354
 Varian, H.R., 352, 353, 357
 Varignon, P., 88
 Vaucanson, J. de, 52, 68
 Velmans, M., 295, 300
 Veuille, M., 362
 Vighetti, P., 361
 Villebressieu, E. de, 51, 66
 Villermé, L.R., 58
 Virilio, P., 67, 68, 70
 Voltaire, 230
 Vuillemin, J., 8, 10, 20, 21, 350, 357
 Vulpian, A., 194
 Vygostky, L., 246, 250

W

Wachsmuth, I., 249
 Waddington, C.H., 17, 37, 38, 45, 49
 Wade, M.J., 206
 Wagner, P., 20, 22, 363
 Wakker, P., 351, 352, 355
 Walliser, B., 365
 Warner, R., 293, 300
 Warrington, E.K., 295, 300
 Wason, P., 233, 297
 Watt, J., 64
 Weinberg, S., 132, 138–140, 149
 Weismann, A., 37, 45, 49
 Weissenbach, J., 68
 Weizkrantz, L., 295
 Weizsäcker, V. von, 38, 39, 46, 49
 Weyl, H., 139, 140
 Whewell, W., 2, 3, 19, 22, 30, 44, 49, 201,
 202, 211, 212
 Whitehead, A.N., 17, 38, 39, 45, 46, 48, 49,
 180, 183
 Wiener, N., 60, 64
 Wiener, O., 310, 351, 354
 Wigner, E.P., 155
 Williams, J.C.P., 248
 Williamson, T., 33
 Wilson, D.S., 141, 206, 300, 365
 Wimsatt, B., 229, 250, 253
 Wittgenstein, L., 9, 21, 22, 44, 49, 96, 101,
 161, 261, 363
 Wolpert, L., 206
 Woolf, A., 64
 Wright, G.H. von, 34, 66, 200
 Wunenburger, J.J., 20, 21
 Wurmser, R., 215
 Wyman, J., 215, 361

Y

Yaari, M., 338, 339, 353, 357

Z

Zahar, E., 110
 Zamir, S., 365
 Zeeman, P., 128
 Zermelo, E., 158
 Zwirn, A., 13, 16, 141, 162, 163, 364, 365