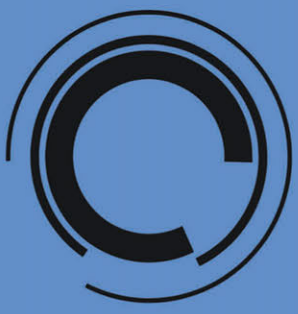
Induction and Deduction in the Sciences

edited by
Friedrich Stadler



Vienna
Circle
Institute
Yearbook

Springer Science+Business Media, B.V.

INDUCTION AND DEDUCTION IN THE SCIENCES

VIENNA CIRCLE INSTITUTE YEARBOOK [2003]

11

VIENNA CIRCLE INSTITUTE YEARBOOK [2003]
11

Institut 'Wiener Kreis'
Society for the Advancement of the Scientific World Conception

Series-Editor:
Friedrich Stadler
University of Vienna, Austria and Director, Institut 'Wiener Kreis'

Advisory Editorial Board:
Rudolf Haller, *University of Graz, Austria, Coordinator*
Martin Carrier, *University of Bielefeld, Germany*
Nancy Cartwright, *London School of Economics, UK*
Robert S. Cohen, *Boston University, USA*
Wilhelm K. Essler, *University of Frankfurt/M., Germany*
Kurt Rudolf Fischer, *University of Vienna, Austria*
Michael Friedman, *Stanford University, USA*
Maria Carla Galavotti, *University of Bologna, Italy*
Peter Galison, *Harvard University, USA*
Adolf Grünbaum, *University of Pittsburgh, USA*
Rainer Hegselmann, *University of Bayreuth, Germany*
Michael Heidelberger, *University of Tübingen, Germany*
Jaakko Hintikka, *Boston University, USA*
Gerald Holton, *Harvard University, USA*
Don Howard, *University of Notre Dame, USA*
Allan S. Janik, *University of Innsbruck, Austria*
Richard Jeffrey, *Princeton University, USA*
Andreas Kamlah, *University of Osnabrück, Germany*
Eckehart Köhler, *University of Vienna, Austria*
Anne J. Kox, *University of Amsterdam, The Netherlands*
Elisabeth Leinfellner, *University of Vienna, Austria*
Werner Leinfellner, *Technical University of Vienna, Austria*
James G. Lennox, *University of Pittsburgh, USA*
Brian McGuinness, *University of Siena, Italy*
Kevin Mulligan, *Université de Genève, Switzerland*
Elisabeth Nemeth, *University of Vienna, Austria*
Julian Nida-Rümelin, *University of Göttingen, Germany*
Helga Nowotny, *ETH Zürich, Switzerland*
Erhard Oeser, *University of Vienna, Austria*
Joëlle Proust, *École Polytechnique CREA Paris, France*
Alan Richardson, *University of British Columbia, CDN*
Peter Schuster, *University of Vienna, Austria*
Jan Šebestík, *CNRS Paris, France*
Karl Sigmund, *University of Vienna, Austria*
Hans Sluga, *University of California at Berkeley, USA*
Elliott Sober, *Stanford University, USA*
Antonia Soulez, *Université de Paris 8, France*
Wolfgang Spohn, *University of Konstanz, Germany*
Christian Thiel, *University of Erlangen, Germany*
Walter Thirring, *University of Vienna, Austria*
Thomas E. Uebel, *University of Manchester, UK*
Georg Winckler, *University of Vienna, Austria*
Ruth Wodak, *University of Vienna, Austria*
Jan Woleński, *Jagiellonian University, Cracow, Poland*
Anton Zeilinger, *University of Vienna, Austria*

Honorary Consulting Editors:
Kurt E. Baier
Francesco Barone †
C.G. Hempel †
Stephan Körner †
Saul A. Kripke
Henk Mulder †
Arne Naess
Paul Neurath †
Willard Van Orman Quine †
Marx W. Wartofsky †

Review Editor:
Michael Stöltzner, *University of Bielefeld, Germany*

Editorial Work/Layout/Production:
Laurence C. Fauth
Hartwig Jobst
Robert Kaller
Camilla R. Nielsen

Editorial Address:
Institut 'Wiener Kreis'
Museumstrasse 5/2/19, A–1070 Wien, Austria
Tel.: +431/5261005 (international)
or 01/5261005 (national)
Fax.: +431/5248859 (international)
or 01/5248859 (national)
email: ivc.zuef@univie.ac.at
homepage: http://univie.ac.at/ivc/

The titles published in this series are listed at the end of this volume.

INDUCTION AND DEDUCTION IN THE SCIENCES

Edited by

FRIEDRICH STADLER
University of Vienna,
and Institute Vienna Circle, Austria



Springer Science+Business Media, B.V.

A C.I.P. Catalogue record for this book is available from the Library of Congress.

ISBN 978-90-481-6555-1 ISBN 978-1-4020-2196-1 (eBook)
DOI 10.1007/978-1-4020-2196-1
Series ISSN 0929-6328

Printed on acid-free paper



Gedruckt mit Förderung der European Science Foundation
Printed with financial support of the European Science Foundation

The European Science Foundation (ESF) acts as a catalyst for the development of science by bringing together leading scientists and funding agencies to debate, plan and implement pan-European scientific and science policy initiatives.
ESF is the European association of 76 major national funding agencies devoted to scientific research in 29 countries. It represents all scientific disciplines: physical and engineering sciences, life and environmental sciences, medical sciences, humanities and social sciences. The Foundation assists its Member Organisations in two main ways. It brings scientists together in its EUROCORES (ESF Collaborative Research Programmes), Scientific Forward Looks, Programmes, Networks, Exploratory Workshops and European Research Conferences to work on topics of common concern including Research Infrastructures. It also conducts the joint studies of issues of strategic importance in European science policy.
It maintains close relations with other scientific institutions within and outside Europe. By its activities, the ESF adds value by cooperation and coordination across national frontiers and endeavours, offers expert scientific advice on strategic issues, and provides the European forum for science.

Gedruckt mit Förderung des Österreichischen Bundesministeriums
für Bildung, Wissenschaft und Kultur
Printed with financial support of the Austrian Ministry for Education, Science and Culture

In cooperation with the *University of Vienna, Center for Interdisciplinary Research / Zentrum für überfakultäre Forschung*

All Rights Reserved
© 2004 Springer Science+Business Media Dordrecht
Originally published by Kluwer Academic Publishers in 2004.
Softcover reprint of the hardcover 1st edition 2004
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.

EDITORIAL

The articles in the general part of this Vienna Circle Institute Yearbook "Induction and Deduction in the Sciences" document the invited contributions to the 2nd international workshop of the European Science Foundation's (ESF) Scientific Network on "Historical and Contemporary Perspectives of Philosophy of Science in Europe". This event, which was organised by the Vienna Circle Institute, took place July 7-9, 2002 at the University of Vienna. The first workshop of this network on main issues, methods and results in history and philosophy of science in an European context was held at the Bertinoro Conference Centre of the University of Bologna from September 30 to October 2, 2001. The network is run by a coordinating committee consisting of Maria Carla Galavotti (Bologna, Italy) who has served as chairperson, Aristides Baltas (Athens, Greece), Donald Gillies (London, UK), Theo Kuipers (Groningen, The Netherlands), Ilkka Niiniluoto (Helsinki, Finland), Michel Paty (Paris, France), Miklos Rédei (Budapest, Hungary), Friedrich Stadler (Vienna, Austria), and Gereon Wolters (Konstanz, Germany). The proceedings of the first workshop have been published in a volume under the title *Observation and Experiment in the Natural and Social Sciences,* edited by Maria Carla Galavotti with Kluwer Academic Publishers.[1]

The third workshop on "Laws and Models in Science" took place at King's College London, September 7-9, 2003 and addressed problems of the structure of science, in particular the role played by theories, laws and models in several scientific fields and their different applications. Its proceedings will be published 2004 in a special volume of the "King's College Publications", edited by Donald Gillies.

The general aim of the ESF network from 2000 to 2003 was to identify the main features of philosophy of science in Europe today, in a spirit of continuity with a tradition that originated in Logical Empiricism (Vienna Circle and Berlin Group) in Central Europe in the first decades of the last century. Since the heyday of scientific philosophy new archival material and research accompanied by extensive transatlantic exchanges has brought forth new perspectives and results in this interdisciplinary field. These contributions have contributed to a clarification and better understanding of the nature of scientific knowledge from a historical and philosophical point of view.[2]

---

[1] Maria Carla Galavotti (ed.), *Observation and Experiment in the Natural and Social Sciences.* Kluwer: Dordrecht-Boston-London 2003. (= Boston Studies in the Philosophy of Science, Vol.232).

[2] Cf. the list of the publications of the Vienna Circle Institute: http://www.univie.ac.at/ivc/

The articles in this volume deal with the main inferential methods that can be applied to different kinds of experimental evidence. These contributions aim at removing the traditional opposition between inductivists and deductivists (e.g., by focusing on abduction) in order to explore the different methods of explanation and justification in the formal, natural and social sciences in different contexts and with different objectives.

This volume also includes critical comments made by leading scholars on methods of the sciences, especially on induction, deduction, abduction, laws, probability and explanation, ranging from logic, mathematics, natural to the social sciences. They present a highly topical pluralist re-evaluation of methodological and foundational procedures and reasoning, e.g. focusing on Bayesianism and Artificial Intelligence.

The workshop in Vienna was funded by the ESF and the organisers, the Vienna Circle Institute together with the University of Vienna, Center for Interdisciplinary Research and Department of Contemporary History. The conference and publication were made possible by the support of the editorial team of Vienna Circle Institute with Gloria Sultano, Daria Mascha, Robert Kaller, Laurence C. Fauth, Camilla R. Nielsen and Hartwig Jobst, all of whom I would like to thank for their professional cooperation. The publication of each Yearbook is supported by the Austrian Ministry of Education, Science and Culture, for which we are grateful as well.

We have retained the format of our series in the 11th volume of the Vienna Circle Yearbook with the announcement of our activities, documentation and a review section presenting recent publications on logical empiricism and related fields, for which I would like to thank our review editor Michael Stöltzner.

We have also included an obituary on Richard Jeffrey, one of the most important and distinguished 20th century philosophers of science who supported IVC from the beginning as a member of the scientific board and contributor to its conferences. We will remember his life and extraordinary work in the tradition of Logical Empiricism in the spirit of his open-minded and creative scholarship.

Vienna, October 2003

Friedrich Stadler
(University of Vienna,
and Vienna Circle Institute)

TABLE OF CONTENTS

A. INDUCTION AND DEDUCTION IN THE SCIENCES

I.

FRIEDRICH STADLER: Induction and Deduction in the Philosophy of Science: a Critical Account since the *Methodenstreit* 1

MALACHI HACOHEN: Historicizing Deduction: Scientific Method, Critical Debate, and the Historian 17

II.

THEO A.F. KUIPERS: Inference to the Best Theory, rather than Inference to the Best Explanation – Kinds of Abduction and Induction 25

ADAM GROBLER: The Significance of Explanatory Considerations 53

III.

ILKKA NIINILUOTO: Truth-seeking by Abduction 57

STATHIS PSILLOS: Inference to the Best Explanation and Bayesianism 83

IV.

JOKE MEHEUS: Adaptive Logics and the Integration of Induction and Deduction 93

MATTI SINTONEN: Argument, Inference and Reasoning – Integrating Induction and Deduction 121

V.

WOLFGANG SPOHN: Laws are Persistent Inductive Schemes 135

ECKEHART KÖHLER: Physical Intuition as Inductive Support 151

VI.

PETER CLARK: Frege, Neo-Logicism and Applied Mathematics 169

MICHEL PATY: Remarks About a "General Science of Reasoning" 185

JEAN-JACQUES SZCZECINIARZ: Two Questions About the Revival of Frege's Programme 195

VII.

DONALD GILLIES: Handling Uncertainty in Artificial Intelligence, and the Bayesian Controversy 199

STEPHAN HARTMANN: Artificial Intelligence and Its Methodological Implications 217

VIII.

HANS ROTT: Supplying Planks for Neurath's Boat: Can Economists Meet the Demands of the Dynamics of Scientific Theories? 225

NILS-ERIC SAHLIN: Informational Economy and Creativity 247

IX.

I. GRATTAN-GUINNESS: The Place of the Notion of Corroboration in Karl Popper's Philosophy of Science 251

LADISLAV KVASZ: How can a Falsified Theory Remain Corroborated? 263

X.

KARL MILFORD: Inductivism in 19th Century German Economics 273

DAVID MILLER: The Uniformity of Nature: What Purpose does it Serve? 293

B. GENERAL PART

REPORT – DOCUMENTATION

SYBILLA NIKOLOW: Planning, Democratization and Popularization with ISOTYPE, ca. 1945: a Study of Otto Neurath's Pictorial Statistics with the Example of Bilston, England 299

REVIEWS

Paolo Parrini, Wesley C. Salmon & Merrilee H. Salmon (eds.), *Logical Empiricism: Historical & Contemporary Perspectives*. Pittsburgh: University of Pittsburgh Press, 2003.
(Christopher Pincock) 331

C. Wayne Mayhall, *On Logical Positivism*, Wadsworth Philosophical Topics, Belmont, 2003.
(Matthias Neuber) 335

C. Wayne Mayhall, *On Carnap*, Wadsworth Philosopher Series, Wadsworth / Thomson Learning, Belmont USA, 2002.
(Christian Damböck) 338

Jean-Pierre Cometti and Kevin Mulligan, *La philosophie autrichienne de Bolzano à Musil. Histoire et actualité*, Paris: J. Vrin, 2001.
(Anastasios Brenner) 339

Steven Beller (ed.), *Rethinking Vienna 1900*. Berghahn Books, New York-Oxford 2001.
(Kurt Rudolf Fischer) 342

Eckart Menzler-Trott, *Gentzens Problem. Mathematische Logik im nationalsozialistischen Deutschland*, Birkhäuser: Basel-Boston-Berlin 2001.
(Matthias Wille) 343

Activities of the Institute Vienna Circle

Activities 2003 347

Preview 2004 350

Obituary

Remembering Dick Jeffrey (1926–2002)
(Maria Carla Galavotti) 353

Index of Names 355

FRIEDRICH STADLER

INDUCTION AND DEDUCTION IN THE PHILOSOPHY OF SCIENCE: A CRITICAL ACCOUNT SINCE THE *METHODENSTREIT*

FROM AN HISTORICAL AND PRAGMATIC POINT OF VIEW

Ever since Aristotle it has been accepted that there exists a combination of inductive and deductive reasoning and a sort of unified inductive-deductive methodology. If one analyzes the procedures and logic of scientific explanation and the methods of generating and justifying scientific knowledge, one recognizes the prototype of philosophy of science found in Aristotle's inductive and deductive procedure that is described in his *Posterior Analytics*, *Physics* and *Metaphysics,* where he

> viewed scientific inquiry as a progression from observations to general principles and back to observation. He maintained that the scientist should induce explanatory principles from the phenomena to be explained, and then deduce statements about the phenomena from premisses which include these principles.[1]

The decisive step taken here is the move from the transition from knowledge of a fact to knowledge of the reasons for the fact. And it is obvious that the general principles of induction, together with logical inference, remain an issue of debate up to this very day. The empirical prerequisites of scientific explanation, the demarcation of empirical from formal science (pure mathematics) as well as the postulated necessary status of the first principles continue to be non-trivial problems in any subsequent philosophy of science after Aristotle.

This history of philosophy of science encompasses the variation and development of Aristotelian methods in the Middle Ages (Roger Bacon, Duns Scotus, William of Ockham et al.), the challenge of the rise of modern science (Galileo, F. Bacon, Descartes) which reformulated scientific laws and led to the controversial debate on inductivism versus the hypothetico-deductive view of science in the 19th century. A special perspective is offered in John Stuart Mill's theory of induction, which explicitly deals with the correlating context of discovery and context of justification. As in later decades, inductivism was linked primarily to the first context and deductivism with the second one. This perspective enables us to examine this unresolved issue against the backdrop of this methodological dualism.

F. Stadler (ed.), Induction and Deduction in the Sciences, 1–15.
© 2004 *Kluwer Academic Publishers.*

Accordingly, I will begin my line of argumentation with this distinction between inductivist and anti-inductivist procedures, identifying the verificationist and falsificationist implications with the aim of clarifying the corresponding epistemological concepts, namely, *subjective* vs. *objective* or *absolute* vs. *relative* knowledge. My surprising conclusion is that there is no contradiction in defending relativism as an objective enterprise which could also be reformulated as a fallibilist methodology.

The fact that this problem is primarily found in the field of the natural sciences raises the question of whether there is an analogous dispute in the cultural and social sciences which could be identified in some variations of the so called *Methodenstreit*. This debate took place in the German speaking world in the late 19th and early the 20th century and which has also become manifest in modern philosophy of science since the Vienna Circle.[2]

After this short preview it is not difficult to imagine that one central figure in this intellectual journey is Karl Popper, who dismissed Aristotle as a minor philosopher. He was probably the first to postulate in the tradition of Hume the theoretical impossibility of each form of induction as a scientific method. The fact that Popper challenged with emotion and knowledge the interplay between induction and deduction was certainly something very new.[3] At the beginning of the 20th century, there was still the coexistence of induction and deduction in philosophy of science, e.g. in Ernst Mach's *Knowledge and Error* (1905)[4], which the author dedicated "to the memory of David Hume, Richard Avenarius and Wilhelm Schuppe". The chapter on deduction and induction from a psychological perspective presents case studies on Mill, Euclid, Bernoulli, which led Mach to the surprising conclusion that "the name 'inductive sciences' for the natural sciences is therefore not justified".[5]

And John Dewey, who signficantly enriched methodology of science with his contribution on the theory of abduction, writes in the chapter on "Scientific Method: Induction and Deduction" in his *Logic. The Theory of Inquiry* (1938):[6]

> Any account of scientific method must be capable of offering a coherent doctrine of the nature of induction and deduction and of their relations to one another, and the doctrine must accord with what takes place in actual scientific procedure.

Decades before the classical writings of Hempel/Oppenheim and Wesley Salmon on the methodology of explanation appeared Herbert Feigl, who advocated epistemological realism throughout his whole life, stated in 1934:[7]

> The purpose of this paper is to make clear (1) that the widely recognized fromulations of the principle of induction do not express the most fundamental rule of induction; (2) that the current view concerning the probability of induction must be revised in terms of a frequency theory of probability; (3) that on this basis the problem of induction in its traditional form is a pseudoproblem; and (4) that the principle of induction must be interpreted as a pragmatic or operational maxim.

And many years later, in a volume dedicated to Rudolf Carnap, Feigl referred to the debate on inductivism/deductivism at the end of his article on "Research Programmes and Induction". Here he took issue with the Popperians and the option suggested by Lakatos:[8]

> In sum: the revised inductivism here submitted provides a justification (vindication) for inductive extrapolations (and interpolations). These inferences not only do occur, but are indispensable for the growth of knowledge. Probabilities for such inferences require background assumptions whose vindication must ultimately be secured in the same manner. There is no other way than the one expressed in the well-known proverbs: 'Look (at least a little) before you leap!' and 'If at first you do not succeed, try, try again'. The sort of vindication sketched above is the one step we can take beyond Hume and skepticism ...

Generally, the dualism of *context of discovery* ("Entstehungszusammenhang") and *context of justification* ("Begründungszusammenhang") can be seen as most closely linked to the classical period of Logical Empiricism. The writings of Rudolf Carnap and Hans Reichenbach are often used to illustrate the manifestation of this crucial methodological distinction. It also appears in Karl Popper's *Logik der Forschung* (1934), whose title, in my view, is misleading especially in the author's translation as *Logic of Scientific Discovery* (1959) instead of, e.g., *Logic of Research* or even *Logic of Justification*. And one might even consider going back to the roots: to Kant's *quid facti* vs. *quid juris* or John Herschel's and William Whewell's explicit treatments of this methodological dualism.

A critical analysis of the influential "received view"-image of the so-called (neo-)positivist philosophy of science[9] could have the following consequences:

- the (inductivist) position of Reichenbach and Carnap could be contextualized in the various expressions of the *Methodenstreit* and against the political backdrop since the turn of the 19th century in order to achieve an adequate judgment of this issue,
- there are good (historical and theoretical) reasons for trying to overcome the dualism by focusing on the remarkable efforts of combining history, sociology *and* philosophy of science in the development of the Vienna Circle / Logical Empiricism (exemplified by the Unity of Science project) before and after World War II,
- this account can be exemplified in addition by the recently contested dualism of relativism and objectivism as a related key concept reaching up to the controversial debates of the so called "Science Wars".

In light of the prevailing image of Logical Empiricism (LE) as an ahistorical or even anti-historical movement, which is primarily associated to the basic dualism of the *context of discovery* (CD) and *context of justification* (CJ) it is at first sight surprising that we find a number of counter-examples. Above all, we have to look at why some members and adherents of the Vienna Circle (VC) or propo-

nents of LE explicitly or implicitly addressed issues related to history and philosophy of history, history of science and history of philosophy of science.[10]

Recent contributions to this rather marginalized field of research (at least as compared to philosophy and methodology of the natural sciences), show – contrary to some of the alleged *Dogmas of LE* (from Quine to Stegmüller) which I have reconstructed and criticized elsewhere[11] – that (1) LE uses syntax and semantics *and* pragmatics, and can thus not be characterized as an ahistorical and formalist/rationalist theoretical school, and (2) that LE does not necessarily represent an inductivist and verificationist school.

If we take a closer look at this dualism of justification and discovery, it is obvious that the early Carnap adhered to the priority of a rational reconstruction as well as to the (methodological) distinction of "Theoretical Questions and Practical Decisions" (1934a)[12] and Popper introduced just this distinction again in his *Logik der Forschung/Logic of Scientific Discovery* (1934/1959).

The most often cited classical reference to the contested dualism of LD/LJ is without doubt Hans Reichenbach's *Experience and Prediction. An Analysis of the Foundations and the Structure of Knowledge* (1938/1983)[13]. But surprisingly, this book, written in Turkish exile, provides only two short passages featuring the distinction as a tool for epistemology in the field of scientific philosophy according the notion of "rational reconstruction" Rudolf Carnap developed in his *Der Logische Aufbau der Welt/Logical Structure* (1928):

> If a more convenient determination of this concept of rational reconstruction is wanted, we might say that it corresponds to the form in which thinking processes are communicated to other persons instead of the form in which they are subjectively performed. The way, for instance, in which a mathematician publishes a new demonstration, or a physicist his logical reasoning in the foundation of a new theory, would almost correspond to our concept of rational reconstruction, and the well-known difference between the thinker's way of presenting it before a public may illustrate the difference in question. I shall introduce the terms context of discovery and context of justification to mark this distinction. Then we have to say that epistemology is occupied in constructing the context of justification.[14]

Referring to this argument he later on concluded:

> At the beginning of our analysis ...we introduced the distinction between the context of discovery and the context of justification. We emphasised that epistemology cannot deal with the first but only with the latter. We showed that the analysis of science is not oriented at the actual processes of thought but at the rational reconstruction of knowledge. We have to remember this definition of epistemology when we want to create a theory of scientific research.[15]

This influential argument recurs in Reichenbach's popular *The Rise of Scientific Philosophy* (1951), which is described in the following way on the dust jacket: "A Psychological Criticism of the Older Systems of Philosophy and a Nontechnical Presentation of the New Philosophy based on Modern Science". Here the decisive passage can be found in the chapter on "Predictive Knowledge" in

which the author defends inductive inference as a method of justification on the basis of observational data:[16]

> The mystical interpretation of the hypothetico-deductive method as an irrational guessing springs from a confusion of context of discovery and context of justification. The act of discovery escapes logical analysis; there are no logical rules in terms of which a 'discovery machine' could be constructed that would take over the creative function of the genius. But it is not the logician's task to account for scientific discoveries; all he can do is to analyze the relation between given facts and a theory presented to him with the claim that it explains these facts. In other words, logic is concerned only with the context of justification. And the justification of a theory in terms of observational data is the subject of the theory of induction.

In this context I would like to take a look at the background motives of Reichenbach who stated right up front in the preface:[17] "If a presentation of philosophy is to be objective, it should therefore be objective in the standards of its critique rather than in the sense of a philosophic relativism." – This formulation also strikes me as being a key passage for the theoretical journey of contemporary philosophy of science. Therefore, I have to come back by clarifying the diametrically opposed concepts of relativism and subjectivism vs. objectivism and rationality. Needless to say, both discovery and justification are intricately linked with each other.

After these two short passages (and some, in part weak, arguments) the question arises why there was such a methodological turn in the philosophy of science in the wake of the (transfer and) transformation of Logical Empiricism after its was forced into exile from continental Europe to Britain and America.[18]

In an attempt to answer this question I will address some characteristic, although not well known, features of this development, which was already anticipated by Carnap himself:[19]

> According to the present view, this theory comprises, in addition to logical syntax, mainly two further fields, i.e., semantics and pragmatics. Whereas syntax is purely formal, i.e., only studies the structure of linguistic expressions, semantics studies the semantic relationship between expressions and objects or concepts; ... Pragmatics also studies the psychological and sociological relations between persons using the language and the expressions.

With the conceptualization of this form of "*Wissenschaftslogik*" (Carnap), which was already in place before the representatives of this philosophy of science were forced to emigrate to the United States, we have also defined the field of philosophy of science as well as the terminological structure for the Unity of Science-movement from 1934 on. It also embodies a philosophy of science from an "pragmatic point of view".[20]

ON RELATIVISM – A CONTROVERSIAL KEY CONCEPT FROM *METHODENSTREIT* TO "SCIENCE WARS"

In the context of the Science Wars debate from 1998 on even certain forms of relativism and irrationalism came under fire in the critique of Alan Sokal and Jean Bricmont which is directed explicitly at Quine, Popper, Kuhn and Feyerabend. Especially in the section entitled "Intermezzo: Epistemic Relativism in the Philosophy of Science", the theoretical physicist Bricmont elaborates on the differentiation of realism/objectivism from all postmodern alternatives implicated in the "Sokal hoax."[21]

I read the following passage about the central notion and theoretical concept of relativism as being entirely in keeping with Larry Laudan's *Science and Relativism* (1990):[22]

> Roughly speaking, we shall use the term 'relativism' to designate any philosophy that claims that the truth or falsity of a statement is relative to an individual or to a social group. One may distinguish different forms of relativism according to the nature of the statement in question: ... Here we shall be concerned only with epistemic relativism and not with moral or aesthetic relativism, which raise very different issues.

Referring to this definition, both authors argue against what they saw as the inevitable consequences of solipsism and radical skepticism, the view of science as practice, the skepticist consequences of a formalized scientific method (according to the Vienna Circle, and others), which led up to a crisis of epistemology. This skepticism can also be found in Popper's *Logic of Scientific Discovery* and its (anti-inductivist) falsificationism, which for his critics (Quine, Kuhn, Feyerabend) was the point of departure for relativist and irrationalist tendencies.

Both the Duhem-Quine-thesis (to which also Otto Neurath contributed significantly) of underdetermination as well as Thomas Kuhn's incommensurability of paradigms do not legitimate a pernicious relativism, since Kuhn believed "*that changes of paradigm are due principally to non-empirical factors and that, once accepted, they condition our perception of the world to such an extent that they can only be confirmed by our subsequent experiences.*"[23]

Finally, it does not come as a surprise that Feyerabend is seen as the intellectual mentor of (methodological) relativism based on the "Copenhagen interpretation of quantum mechanics", even though not of irrationalism, through his "anything goes" recommendation:[24]

> One could defend these assertions by invoking the classical distinction between the context of discovery and the context of justification. Indeed, in the idiosyncratic process of inventing scientific theories, all methods are in principle admissible – deduction, induction, analogy, intuition and even hallucination – and the only real criterion is pragmatic. On the other hand, the justification of theories must be rational, even if this rationality

cannot be definitively codified. One might be tempted to think that Feyerabends's admittedly extreme examples concern solely the context of discovery, and that there is thus no real contradiction between his viewpoint and ours.

But the problem is that Feyerabend explicitly denies the validity of the distinction between discovery and justification. Of course, the sharpness of this distinction was greatly exaggerated in traditional epistemology. We always come back to the same problem: it is naive to believe that there exist general, context-independent rules that allow us to verify or falsify a theory; otherwise put, the context of justification and the context of discovery evolve historically in parallel. Nevertheless, at each moment of history, such a distinction exists. If it didn't, the justification of theories would be unconstrained by any considerations of rationality.

Here we should remember that precisely this issue of the controversial epistemological relativism – as a sort of threat to science and society – was already the subject of heated debate among Otto Neurath, Philipp Frank, Popper and others between the World Wars I and II.[25] And a similar dispute can be reconstructed in the exchanges between main members of the Vienna Circle and the Austrian School of Economics, which became manifest in British exile contacts between Karl Popper, Friedrich August von Hayek and Otto Neurath.[26]

Let me return to my previous considerations:

I would *first* like to argue that a more precise historical study and analysis of the development of European philosophy of science over the past hundred years beginning with the *Methodenstreit* (methodological debate) in the German political economy and leading up to the present gives us a chance to reconfigure clearly polarized positions by pointing to internal alternatives.

My *second* claim is that the above-mentioned problems refer to unsolved problems on the one hand and actually irreconciliable methodological and philosophical positions on the other. A quick glance at the various manifestations of the *Methodenstreit* up to the present prompts some thought experiments which may lead us further. I would like to focus mainly on the following debates in the twentieth century:

- the classical controvery between the German historical school (Gustav Schmoller) and the Austrian School in economics (Carl Menger)
- the central dispute in the historical sciences (associated with Karl Lamprecht)
- the debate on value statements in the cultural sciences (associated with Max Weber)
- three variations on the so-called "positivism dispute" with Lenin vs. Ernst Mach, Max Horkheimer vs. Otto Neurath, Karl Popper/Hans Albert vs. Theodor W. Adorno/Jürgen Habermas)
- the dispute on the sociology of knowledge (around Max Scheler and Karl Mannheim)
- the continuing controversy between the pure theory of law (Hans Kelsen) and natural law (Carl Schmitt)

The division in up to three science cultures (of *Geisteswissenschaften*, natural sciences, social sciences) is also the product of different answers situated between induction and deduction – answers to explanation vs. understanding associated with the "narrative turn", to the unity or diversity regarding the subject and the method of science, to the question of foundational knowledge. We can identify different types of philosophy of science that are not only embedded in a local context but also have an intercultural meaning. Thus the question as to the contextuality of knowledge was already fully articulated in the 1920s in the debate on the sociology of knowledge, social epistemology and social science studies.

The postulate of an objective "real" external world independent of all subjects of knowledge was already Lenin's main objection against Mach's relativism qua fallibilism presented in his *Knowledge and Error. Sketches on the Psychology of Enquiry* (1905/1976).[27]

And contemporary scientists such as Steven Weinberg (as opposed to the "Kuhnians") advocate this contrary pair of epistemological relativism and all forms of realism/objectivism/rationalism. The scientific character of history as a theoretical discipline was at the center of a long debate in the German political economy between Gustav Schmoller and Carl Menger. In his *The Logic of Scientific Discovery* and *Poverty of Historicism* Karl Popper used the controversial distinction between discovery and justification of scientific theories to offer a classical alternative to this dualism: namely, the *methodological* unity of all sciences based on the hypothetical theoreticity of knowledge in all disciplines – but, surprisingly, excluding history and historiography. His "critical rationalism" has hardly delineated itself from the (methodological) dualism of the natural and social sciences in Friedrich August von Hayek's philosophy of science (especially his critique of "scientism" as a paradigm of the natural sciences). However, both conceptions were already called into question by Otto Neurath, who described Popper's position as "Pseudorationalism of Falsification" (1935).[28] His main objection was the latter's assumption of an idealized science ("Realwissenschaft") with a consistent rationality serving as a clear criterion for rejecting and accepting of theories. According to Neurath's idea of an empiricist encyclopedia we should see the sciences metaphorically as a boat that is constantly being rebuilt as part of a collective effort and does not justify any absolute position from the outside. Accordingly, one version reads:[29]

> We are like sailors who on the open sea must reconstruct their ship but are never able to start afresh from the bottom. Where a beam is taken away a new one must at once be put there, and for this the rest of the ship is used as support. In this way, by using the old beams and driftwood the ship can be shaped entirely anew, but only by gradual reconstruction.

This sort of methodological holism and non-reductive naturalism included a rejection of all ultimate foundations of science and the unambiguity of knowledge by means of an all-decisive criterion of rationality.[30] This has become

generally accepted as a result of Quine's writings in the American scientific community. Paradoxically, it has reanimated the *received view* in the philosophy of science that was directed against Logical Empiricism.

As a result of the local 'cultural and science wars' in the USA and France, an irony of the history of science went unnoticed by its supporters. While Sokal / Bricmont and their proponents in philosophy of science, e.g., Noretta Koertge (1998)[31] mainly given "relativism" (together with "constructivism") the blame for all irrational and decisionist ills, the very development of the philosophy of science in, and also since, the Vienna Circle has broken with the anathema of Cartesian certism and moved towards a (non-foundationalist) epistemology and (methodological) relativism.[32]

In an attempt to relate theory to observation the Einstein biographer Philipp Frank, for instance, who was sensitized by the cultural resistance to Einstein's theory of relativity in Europe, consistently underscored the conventionalist and constructive character of philosophy of science (by referring to Mach, Poincaré, Duhem, Bridgman, Frege, and others). He thus succeeded in creating a promising 'link between science and philosophy'.[33]

In this context, one of Frank's late publications in American exile is enlightening. In his booklet *Relativity – A Richer Truth*, for which Albert Einstein wrote the preface[34] the leftist intellectual defended himself against the general reproach that European relativism with its agnosticism and skepticism was responsible for the emergence of fascism and national socialism. This is a theme that continued to be a central focus for a number of years, beginning with the "Conference on Science, Philosophy and Religion" in 1940 in New York City. Frank's contributions to this conference were summarized in the book mentioned above, where he sought

> to show that so-called relativism has not the slightest thing to do with agnosticism or skepticism, that it is in no way hostile to the belief in ethical and democratic values, that it is accompanying every advance in science and is nothing but a significant representation of the enrichment of human expression which is inseparably connected with our gradually increasing experience.[35]

And later he wrote more explicitly on the central notion of this cultural struggle:[36]

> The more we discourage the scientific attitude, which is based on 'relativism', the more we run the risk that the ethical or political system which we claim to be 'absolutely true' can be superseded in one rush by the opposite creed which other people believe be 'absolute true'.
>
> My point is: On the one hand the doctrine of the 'relativity of truth' is tied up inseparably with the progress of human knowledge. But on the other hand this doctrine does not imperil by any means the 'objectivity' of truth and does not account for the shrinking belief in 'objectivity of values'.

I would argue that Frank's approach can be directly linked with the main conceptions of the *Methodenstreit* and 'Science Wars' that are now subsiding. In the tradition of Mach, Frank did not question the existence of a real, objective external world or the possibility of a rational reconstruction of theory dynamics in the form of a "thematic analysis of science."[37] What he did do, however, was call into question the idea of a direct, exclusive access to knowledge without any critique of language, theorizing, models of thinking and metatheoretical reasoning. In addressing the 'text' (qua theories) the "context" was always implicitly addressed as well (qua "republic of scholars"), which leads to a correlative refutation of any *experimentum crucis*. Since 1900 this development was also closely linked to the "pragmatic turn" (preceding the "linguistic turn") in the philosophy of science. Therefore, Frank's reference to John Dewey's *Reconstruction in Philosophy,* (1920/1948)[38] is not by chance: "*the author points out that the traditional systems of metaphysics and philosophy do not reveal any truth about the nature of the universe, but rather present the aspirations in metaphorical language.*"[39]

In addition, one should take into account here of the integration of Dewey's philosophy of science in the publication project *International Encyclopedia of Unified Science* 1938ff. (eds. Carnap/Morris, Neurath), especially with his contribution on "Theory of Valuation" (1951).[40]

Why did I mention this particular historical episode? *First*, because it calls into question the clichéd notion of ahistorical "(logical) positivism", and *second*, because it is symptomatic for the current debates between the self-proclaimed relativists and objectivists, or in an recent terminology between advocates of the "realist-rationalist cluster" vs. the "socio-historical cluster".[41] The context of these developments cannot be overseen: after World War II it was without doubt the Cold War accusations against all socialist, liberal and relativist images in science and society that played a crucial role.[42]

Apparently, in the ongoing scientific discussions the anti-Cartesian understanding of science (around "relativism") has generally been made the scapegoat for all ills like "anti-science". Despite well-meaning attempts to salvage the realist-rationalist cluster, this argument has proven to be a non-starter. What makes things even more confusing is the fact that the advocates of an enlightened modernism are giving the academic left and radical feminists part of the blame for the crisis of scientific reason. While this might seem understandable in view of popular science authors such as John Horgan (1996) and some extreme antirationalist positions in feminist studies, it ultimately stands in the way of a productive discussion on the "limits of scientific growth and rationality".

When the Nobel Prize laureate Steven Weinberg denounced every form of "epistemological relativism" along with Kuhn's *Structure of Scientific Revolutions*,[43] he was arguing purely from the position of a hard scientist, oblivious to the most recent discussions in philosophy of science. (On the other hand, I doubt that Steven Fuller's polemic against Kuhn as a "cold warrior" actually helps clear up the ideological muddle).[44]

Neither the "conditions of the possibility of (physical) knowledge" nor the critique of the empirical concept in microphysics can be overcome by a naïve materialism/realism or a postulated objectivism of experimental or theoretical science.

Precisely *in this sense*, Sokal's and Bricmont's attack on the "abuse of science by postmodern scientists" must to be viewed critically, as it does not (adequately) take into account the state of the art in the history of philosophy of science.[45] This is something that was already correctly noted by the journalist Jim Holt in his review of this book[46]: as physicists they succeeded in revealing the pseudo-scientific rhetoric, but as philosophers and philosophers of science they still have much to learn.[47]

To elucidate some of the issues sketched here I would like to ask the following question: do the sciences really represent "a house built on sand" to borrow the title of the anthology directed against postmodern science?[48] Alluding to the critical-rationalist spirit of Popper's thought, this book takes issue with so-called "Postmodern Myths about Science." But alone the title provokes questions, when we take a look at Popper's *Logic of Scientific Discovery*, where it is explicitly stated that there is no absolute empirical basis for the objective sciences, only a sort of swampy land.[49] And surprisingly Popper quotes (inconsistently, I believe) Hermann Weyl's *Philosophy of Mathematics and Natural Science* who – similar to the cited Neurath boat metaphor – employs two diametrically opposed epistemological positions, namely: *subjective-absolute* on the one hand, and *objective-relative* on the other.

> The empirical basis of objective science has thus nothing 'absolute' about it. Science does not rest upon solid bedrock. The bold structure of its theories rises, as it were, above a swamp. It is like building erected on piles. The piles are driven down from above into the swamp, but not down to any natural or 'given' base; and if we stop driving the piles deeper, it is not because we have reached firm ground. We simply stop when we are satisfied that the piles are firm enough to carry the structure, at least for the time being.[50]

CONCLUDING REMARKS

An epistemological alternative to a controversial *relativism* appears to be rather a philosophical *absolutism* (that could not be fruitful to begin with) but not an *objectivism* – if it does not raise obsolete pseudo-questions as to the existence of a real external world. And a possible replacement for the contested concept of relativism would be a pragmatically understood, methodologically formulated *fallibilism* – as the smallest common denominator for a philosophy of science and epistemology that does not attempt to be scientistic or foundationalist. This might be a way to initiate a fruitful discussion after the 'science wars' on a theoretical and methodological level – beyond purely academic power politics – where central questions of present-day research could be addressed. The history

of philosophy and of the sciences would then not be the handmaid of a normative philosophy of science but rather an integral part, a key to understanding the intellectual life. Perhaps an incentive for this could come from the European history and philosophy of science since the Fin-de-Siecle around 1900 with the rich insights it has been able to gain from a more than century old *Methodenstreit.* And this is another attempt to mutually take into account the context of discovery with the context of justification and as a corollary the inductive-deductive reasoning.[51]

NOTES

1. Losee 2001, p. 5
2. Cf. Felix Kaufmann 1936/1999, esp. pp. 133-322.
3. Regarding esp. Karl Popper's philosophy of science in the context of *Methodenstreit* cf. Hacohen 2000 and his commentary in this volume.
4. Mach 1905/1976.
5. Ibid., p. 231.
6. Dewey 1938, p. 419.
7. Feigl 1934, p. 153.
8. Feigl 1971, p. 149f.
9. Stadler (ed.) 2003.
10. Heidelberger/Stadler (eds.) 2002 and 2003.
11. Stadler 2001, esp. Ch. 0.
12. Carnap 1934a. Originally published as "Theoretische Fragen und praktische Entscheidungen".
13. Reichenbach 1938. *Experience and Prediction: An Analysis of the Foundation and the Structure of Knowledge*, Univ. of Chicago, 1957, pp. 6-7 (1st ed. 1938).
14. Reichenbach 1938.
15. Ibid., p. 239.
16. Reichenbach 1951, p. 231.
17. Ibid., p. IX.
18. Stadler 2003a, 2003b.
19. Carnap 1934b, quoted after Carnap 1968, p. VII.
20. Stadler 2003.
21. Sokal/Bricmont 1998.
22. Ibid., p. 51f.
23. Ibid., p. 75.
24. Ibid., p. 81f.
25. Frank 1932.
26. A survey in Stadler 2003a, p. 169ff. Cf. also Hacohen 2000 on the relation of Popper and Hayek.
27. Mach 1905. Lenin 1909.
28. Neurath 1935.
29. Neurath 1921, quoted after Neurath 1973, p. 199.
30. Uebel 2000.
31. Koertge (ed.) 1998.
32. Regarding verificationism a normal science methodology: Mišak 1995.
33. Frank 1949.
34. Frank 1950. Expanded German edition Frank 1952.
35. Frank 1950, p. XV.
36. Ibid., p. 6f.

37. Holton 1994.
38. Dewey 1920.
39. Frank, op.cit. p. 137.
40. Carnap/Morris/Neurath (eds.)1938ff. Reprint 1970f. Dewey 1951.
41. Kitcher 1998 and 2001.
42. Recent research on Cold War context and based on archives is provided by George Reisch, e.g., in: Hardcastele/Richardson 2003.
43. Weinberg 1998.
44. Fuller 2000.
45. Therefore, it is really not surprising that there is a broad epistemological overlapping in the books of Sokal/Bricmont (1998) and Lenin (1909), which caused a rather strange controversy in the e-mail list of the "History of Philosophy of Science Working Group" (HOPOS) at the end of August 2003.
46. Holt 1998.
47. Cf. Kieseppä/Koertge/Stadler 1999.
48. Koertge (ed.) 1998.
49. Popper 1972, Chapter V, §30.
50. Weyl quoted after Popper 1972, p. 111.
51. Cf. the contributions in: Heidelberger/Stadler (eds.) 2002 und 2003.

REFERENCES

Carnap, Rudolf 1934a "Theoretische Fragen und praktische Entscheidungen", in: *Natur und Geist* Nr. 9, pp. 257-260.

Carnap 1934b. *Logische Syntax der Sprache*. Wien: Springer. Second edition 1968. English edition: London: Routledge and Kegan Paul 1937.

Carnap, Rudolf/Charles Morris/Neurath, Otto (eds.) 1938ff. Collected reprint as: *Foundations of the Unity of Science. Toward an International Encyclopedia of Unified Science*. 2 Vlms. Chicago and London: Chicago University Press 1970/1971.

Dewey, John 1920. *Reconstruction in Philosophy*. New York: Henry Holt & Co. Reprint: Boston: Beacon Press 1924.

Dewey, John 1938/2002. *Logic: The Theory of Inquiry*. Carbondale: Southern Illinois University Press. German edition: *Die Theorie der Forschung*. Aus dem Englischen von Martin Suhr. Frankfurt/M.: Suhrkamp.

Dewey, John 1951. *Theory of Valuation*. The University of Chicago Press. (= International Encyclopedia of Unified Science II/4). Reprint in: Carnap/Morris/Neurath 1970f., pp. 379-448.

Feigl, Herbert 1934. "The Logical Character of the Principle of Induction", in: Feigl 1981, pp. 153-163.

Feigl, Herbert, 1971. "Research Programmes and Induction", in: R.C.. Buck/R.S. Cohen (eds.), *PSA 1970. Boston Studies in the Philosophy of Science* VIII, pp. 147-150.

Feigl, Herbert 1981. *Inquiries and Provocations. Selected Writings, 1929-1974*. Dordrecht-Boston-London. Reidel.

Frank, Philipp 1932. *Das Kausalgesetz und seine Grenzen*. Wien: Springer. English edition: *The Law of Causality and its Limits*. Ed. by Robert S. Cohen. Dordrecht-Boston-London: Kluwer 1998.

Frank, Philipp 1949. *Modern Science and its Philosophy*. Harvard University Press.

Frank, Philipp 1950. *Relativity – A Richer Truth*. Preface by Albert Einstein. Boston: Beacon Press. German edition: Zürich: Pan Verlag 1952.

Fuller, Steven 2000. *Thomas Kuhn. A Philosophical History for our Time*. University of Chicago Press.

Hacohen, Malachi Haim 2000. *Karl Popper. The Formative Years, 1902 – 1945. Politics and Philosophy in Interwar Vienna*. Cambridge University Press.

Heidelberger, Michael and Stadler, Friedrich (eds.) 2002. *History of Philosophy of Science. New Trends and Perspectives.* Dordrecht-Boston-London: Kluwer

Heidelberger, Michael and Stadler, Friedrich (eds.) 2003. *Wissenschaftsphilosophie und Politik. Philosophy of Science and Politics.* Wien-New York: Springer.

Holt, Jim 1998. "Is Paris Kidding?", in: *The New York Times Book Review*, Nov. 15, 8.

Holton, Gerald 1994. *Thematic Origins of Scientific Thought. Kepler to Einstein.* Revised Edition. Harvard University Press.

Holton, Gerald 1994. *Science and Anti-Science.* Harvard University Press.

Horgan, John 1996. *The End of Science.* Reading: Addison-Wesley.

Kitcher, Philip 1998. "A Plea for Science Studies", in: Koertge (ed.), 32-56.

Kitcher, Philip 2001. *Science, Truth, and Democracy.* Oxford University Press.

Koertge, Noretta (ed.) 1998. *A House Built on Sand. Exposing Postmodernist Myths about Science.* Oxford University Press.

Koertge, Noretta/Kieseppä, Ilkka/Stadler, Friedrich 1999. "A House Built on Sand? A Conversation with Noretta Koertge", in: Daniel Greenberger/Wolfgang Reiter/Anton Zeilinger (eds.), *Epistemological and Experimental Perspectives on Quantum Physics.* Dordrecht-Boston-London, pp. 279-301.

Kaufmann, Felix 1936/1999. *Methodenlehre der Sozialwissenschaften.* Wien: Springer. Reprint, neu hrsg. von Günther Winkler.

Lenin, V.I. 1909. *Materialismus und Empiriokritizismus. Kritische Bemerkungen über eine reaktionäre Philosophie.* Moskau. English edition: *Materialism and Empirio-Criticism. Critical Notes Concerning a Reactionary Philosophy.* New York: International Publishers 1927.

Losee, John 2001. *A Historical Introduction to the Philosophy of Science.* Fourth edition. New York: Oxford University Press.

Mach, Ernst 1905. *Erkenntnis und Irrtum. Skizzen zur Psychologie der Forschung.* Leipzig: J.A. Barth. English: *Knowledge and Error. Sketches on the Psychology of Enquiry.* With an Introduction by Erwin N. Hiebert. Dordrecht-Boston 1976.

Mišak, C.J. 1995. *Verificationism. Its History and Prospects.* London-New York: Routledge.

Neurath, Otto 1921. *Anti-Spengler.* München: Callwey. Abridged English translation in: Neurath 1973, pp. 158-213.

Neurath, Otto 1935. "Pseudorationalismus der Falsifikation", in: *Erkenntnis* V, 353-365. Reprint in: Neurath 1983, pp. 121-131.

Neurath, Otto 1973. *Empiricism and Sociology.* Ed by M. Neurath and R.S. Cohen. Dordrecht-Boston: Reidel.

Neurath, Otto 1983. *Philosophical Papers 1913-1946.* Ed. by Robert S. Cohen and Marie Neurath. Dordrecht-Boston-Lancaster: Reidel.

Popper, Karl 1934/1972. *Logik der Forschung. Zur Erkenntnistheorie der modernen Naturwissenschaft.* Wien: Springer. First English edition: *The Logic of Scientific Discovery.* London: Hutchinson 1959. (Sixth revised impression 1972).

Reichenbach, Hans 1938. *Experience and Prediction. An Analysis of the Foundations and the Structure of Knowledge.* The University of Chicago Press 1938. Deutsch: *Erfahrung und Prognose. Eine Analyse der Grundlagen und der Struktur der Erkenntnis.* Mit Erläuterungen von Alberto Coffa. Braunschweig-Wiesbaden: Vieweg 1983.

Reichenbach, Hans 1951. *The Rise of Scientific Philosophy.* University of California Press. Deutsch: *Der Aufstieg der wissenschaftlichen Philosophie.* Berlin:Grunewald: Herbig 1951.

Snow, C.P. 1959. *The Two Cultures and a Second Look. An Expanded Version of the two Cultures and the Scientific Revolution.* Cambridge University Press 1964.

Sokal, Alan 1996. "Transgressing the Boundaries: Toward a Transformative hermeneutics of Quantum Gravity", in: *Social Text* 46/47, 217-252.

Sokal, Alan and Bricmont Jean 1998. *Fashionable Nonsense. Postmodern Intellectuals' Abuse of Science.* New York: Picador. French edition: *Impostures Intellectuelles.* Paris: Editions Odile Jacob. German edition: *Eleganter Unsinn. Wie Denker der Postmoderne die Wissenschaften mißbrauchen.* München: Beck 1999.

Stadler, Friedrich 2001. *The Vienna Circle. Studies in the Origins, Development, and Influence of Logical Empiricism.* Wien-New York: Springer.

Stadler, Friedrich (ed.) 2003. *The Vienna Circle and Logical Empiricism. Re-evaluation and Future Perspectives*. Dordrecht-Boston-London: Kluwer.

Stadler, Friedrich 2003a. "The Wiener Kreis in Great Britain: Emigration and Interaction in the Philosophy of Science", in: Timms, Edward and Hughes, Jon (eds.), *Intellectual Migration and Cultural Transformation. Refugees from National Socialism in the English-Speaking World*. Wien-New York: Springer, pp. 155-180.

Stadler, Friedrich 2003b. "Transfer and Transformation of Logical Empiricism: Quantitative and Qualitative Aspects", in: Gary Hardcastle and Alan Richardson (eds.), *Logical Empiricism in North America*. University of Minnesota Press. (forthcoming)

Uebel, Thomas 2000. *Vernunftkritik und Wissenschaft. Otto Neurath und der Erste Wiener Kreis*. Wien-New York: Springer.

Weinberg, Steven 1998. "The Revolution That Didn't Happen", in: *The New York Review of Books* 8/98, 48-52.

Vienna Circle Institute,
and University of Vienna
Museumstraße 5/2/19
A-1070 Vienna
Austria
Friedrich.Stadler@univie.ac.at

MALACHI HACOHEN

HISTORICIZING DEDUCTION: SCIENTIFIC METHOD, CRITICAL DEBATE, AND THE HISTORIAN

Comments on Friedrich Stadler's "Induction and Deduction in the Philosophy of Science"

If ever there were scientific procedures that seemed immune to history, induction and deduction would be them. Their validity seemingly unimpinged by the vicissitudes of history, they appear a proper subject of discussion for philosophers and scientists, but not for historians. Historians pride themselves on demonstrating that the internal logic of theory is historical – a response to particular conditions. Breakdowns in logic present historians with opportune moments for historicization, for showing how theoreticians' efforts to respond to their situation made them move inconsistently. When it comes to induction and deduction, however, they seem so universal and basic – the logic and methodology of science unthinkable without them – that any effort to demonstrate their historically bound character risks relativizing science and, some would argue, undermine rational exchange. Can the historian participate, qua historian, in the conversation on induction and deduction without unraveling science and undermining rational exchange?

Stadler remains unconcerned by this dilemma. He thinks that it is a pseudo-problem, reflecting mistaken assumptions about the incompatibility of relativism and objectivity. He strives to show that the Vienna Circle already addressed the issues dividing scientists and science studies critics. Circle members were able to arbitrate, he assures us, between history and sociology on the one hand and scientific claims to autonomy on the other. Critics misunderstand them because of problems in the transmission of interwar Central European philosophy to the Anglo-Saxon academy. Only misunderstanding could have created the absurd situation whereby science studies critics attack the Logical Empiricists for their objectivism, and the scientists for their relativism. He, the historian, consulting a century of methodological debates (*Methodenstreit*), would show the Science Wars of the 1990s for what they were: a disappointing déjà-vu.

Stadler spares no effort to save the Vienna Circle from the charge of having lived in a netherworld of syntax and semantics, or for having limited their philosophy to rational reconstruction of scientific procedures. They were attentive to pragmatics as well, he argues, and debated scientific practice as a historical and sociological phenomenon. Stadler positions both himself and the circle as moderate and wise peace-makers in a world torn by Science Wars. History –

F. Stadler (ed.), Induction and Deduction in the Sciences, 17–23.
© 2004 *Kluwer Academic Publishers.*

Stadler's history – would make it possible to negotiate peace. The goal is a fallibilist science where history calls attention to science's pragmatic formation, makes those engage – scientists and philosophers alike – better informed, wiser, and more moderate, but does not undermine science or rational discourse.

I sympathize with Stadler's project and share his goal, but I am not sure that he is an effective negotiator. His logical positivist heroes provide far less instruction on the quandary of historicization than he thinks. To be sure, Carnap and Reichenbach sensed the fragility of the scientific method, but their response was to safeguard scientific autonomy by distinguishing between (irrational) discovery and (rational) justification and insisting that justification was amenable to rational reconstruction. Carnap acknowledged the existence of "Pragmatics, [the study of] the psychological and sociological relations between persons using the language [of science]," but his gesture did not translate into systematic exploration of the conditions facilitating rational exchange.[1] Neurath marveled at the political and ideological character of methodological debates, but his commitment to a universal scientific language, grounded in a radical enlightenment view, saved him from relativism. He hoped to build a cosmopolitan scientific community that, imposing draconian linguistic constraints, will assure orderly reconstruction of the boat of science in midsea. Neurath will not save Stadler because precisely the Enlightenment's project and the idea of a universal language are presently under attack. It is comforting to know that theoreticians in different disciplines have been living with relativism for over a century, but it is not altogether clear how Stadler intends to mobilize the resources offered by his survey of past approaches to negotiate theoretically between them. Stadler is in greater trouble than he thinks.

Popper seems to me less vulnerable than the logical positivists to the relativist challenge: He made short shrift of efforts to construct a universal scientific language and declared method itself a conventional set of "rules of the game," open to criticism and change. He disposed of induction on logical grounds. Should deduction prove to have limited value as a methodological tool, the scientific vision may yet be saved, as long as critical discourse, however undisciplined, remained a potential arbiter of disputes. But even to retain this modest vision of scientific rationality, one must uphold rational criticism as a real possibility. Radical historicization seems to challenge this possibility, or, at least, science critics seem to think it does. How can one historicize logical procedures without undermining rational exchange? In the rest of this comment, I shall offer an example of historicization that helps the methodological conversation rather than renders it irrelevant. My theme is: the changing place of deduction in Popper's philosophy.

Popper began his epistemological revolution by disposing of induction and offering an empirico-deductive model of science. In much of the first volume of *Die beiden Grundprobleme der Erkenntnistheorie,* written between 1930 and 1932, he criticized positivist solutions to the problem of induction.[2] To Hume's problem – there is no logically valid progression from particular experiences to

general laws – Popper offered a radical solution: Such progression was unnecessary for science and played no role in it. Deductive models made universal theories (or natural laws) logically tenable by testing single reality statements (*Wirklichkeitsaussage*) deduced from the theory against reality. Potential falsification rendered universal theories, which could not be verified because of the induction problem, partially decidable: They could be falsified in a test, hence they were a part of science. Popper was still a foundationist and a "justificationist" – he believed that single statements could be verified and justify holding a theory. The distinctive mark of scientific statements was their decidability, not their falsifiability. He did not distinguish yet between logical and methodological procedures. Induction was a logical problem, not a methodological one. Science retained affinity with certainty: Scientific decision was a matter of correct logical procedure, not methodological convention. Popper was mighty proud of having surmounted science's foremost logical problem – induction.

Within two years, Popper became a non-foundationist and transformed deductivism from a logical model into a conventional methodology. Falsification, and, later, falsifiability (potential falsification), became the cornerstone of deductivist methodology and the demarcation criterion for science. Statements that could not be tested and falsified were not scientific. In the revised section 11 of volume I of *Grundprobleme,* written during the fall of 1932, and in fragments of the second volume written about the same time, Popper recognized that even basic statements (*Basissätze*) were low-level theories, verification of which could never reach a closure. Their testing did not always produce incontrovertible results and could continue, in principle, *ad infinitum.* A decision to stop testing and accept a statement was conventional. Tests operated by consensus and convention.[3] Method provided ground rules for testing and evaluating scientific statements, and Popper devoted *Logik der Forschung* (1935) to developing deductivism as a methodology. But, even for those accepting the rules, the room for debate and disagreement was wide open. Negotiation over basic test statements became the paradigmatic scientific activity. Scientific evaluations were informed by logical and methodological analysis, but they were neither logically nor methodologically decidable. The possibility of falsification, of the scientific community agreeing that a theory failed a test and another must be substituted, still held the prospect of progressive error elimination. But already at this stage, falsification had as much to do with critical debate, unbound by methodological rules, as it did with deductive procedures. The control exercised by the deductive method loosened.

The shift from method to public debate redefined scientific objectivity. It was no longer a matter of correct procedure, but of intersubjectivity – subjecting all results to public criticism. The full implications were not yet clear in *Logik der Forschung*: It was a book on scientific method, not on public debate, and Popper focused on methodological procedure, not on institutional processes. He presupposed a vital public, but he seemed to take the Central European community of scientists and philosophers, exemplified in the Vienna Circle, for granted: It was

there, and it did not need to be theorized. Fascism destroyed this community soon after, and *The Open Society and Its Enemies* (1945), written in New Zealand during World War II, contemplated the ramifications of the catastrophe. As Popper moved to political philosophy, he recognized similarities between his scientific, moral, and political ideals and confronted the public character of science. Critical debate, he argued, was essential to democracy – a public sphere to debate and negotiate contending policies. He found this sphere exemplified in the Republic of Science – a free cosmopolitan community, engaged in critical debates. He made intersubjective criticism the demarcation criterion for science and democratic politics alike.[4] If, previously, scientific testing was the prime site of objectivity (intersubjectivity), now it appeared as mere instantiation of critical debate, more formal and procedure-bound than political discussion, but not qualitatively different.

In a universe encompassing science and politics, deductive method assumed a marginal role. Political discourse could not operate by deductive rules, else, too much of it (*The Open Society* included) would have had to be ruled irrational. Public debate must therefore be rational without strictly following deductive reasoning. But how? How can political discussion remain rational even when method can no longer offer guidelines? It seems that, when methodological control loosens, in the transition from scientific to political debate, the controls of the (democratic) public sphere kick in. The public character of the debate, the mere articulation of opposing positions in public, is sufficient to sustain rationality. (This is true especially under the conditions of freedom and equality but even when they are imperfect.) Popper seemed to acknowledge democratic controls – democracy provided effective mechanisms for changing rulers, resolving conflicts, and persuading opponents – but he remained virtually silent about their relationship to methodological ones. Indeed, he implied the identity of deductive reasoning, criticism, and rationality. Always aiming at the highest degree of methodological control (which raised precision and clarity), he never considered the implications of discourse unregulated by them, although he showed the existence of such discourse by drawing the limits of method. Important scientific disputes, such as those concerning divergent estimates of the degree of corroboration of competing theories, did not admit of a high degree of methodological control. How much truer was this of politics!

Neither Popper nor his students ever drew full conclusions from the limits of method, or their discussions would have become less focused on it. Rather, they would have raised questions relating to optimal democratic controls, and this would have taken them to discourse ethics and to the sociology of knowledge.[5] One of Popper's major achievements was to lay out the preconditions of rational political discourse and to have shown that they entail several rules of conduct in debate, but he did not develop them fully. Moreover, he barely broached the question of the institutional arrangements most conducive to critical debate. He showed that the end of freedom would spell also the end of science, but he proceeded no further. This left him open to the charge that he proffered an idealized

vision of science and politics. Critical Rationalists must respond to the challenge by considering the institutional controls on the universe of knowledge and their possible reform.

Critical Rationalists consider, however, my distinction between deductive reasoning (or method) and critical debate problematic. They are inclined to think of them as one and the same. Some suggest that arguments often have the appearance of being non-deductive because they presuppose unarticulated background knowledge, and part of rational investigation consists of bringing to light these presumptions, illuminating a proposal's consequences. Political activists, too, can improve their proposals, says Karl Milford, if they evaluate them deductively.[6] I agree that criticism of proposals in its sharpest and clearest form follows deductive reasoning, but such a level of clarity and simplicity, while desirable, is rarely attainable, and, even in its absence, debate does not lose its rationality.[7] The view that it does not only flies in the face of most political, historiographical, and even scientific debate, but also identifies reasoning with method and logical procedure – logicism. If, as Miller argues, "our thinking consists of exploratory problem solving that is not constrained by rules ... [and] real critical thinking ought to be the exposure not of poor arguments but of poor guesses," then insightful counter-proposals, still wild guesses not subject to critical control, or not yet clearly formulated, may well expose the poverty of a competing proposal.[8] Indeed, this is often the way successful political proposals are articulated. Deductive investigation is part of a larger universe of critical discourse. It does not absorb it.

Popper restored unity to the universe of knowledge by demystifying science and arguing that scientific method is but a more formal, systematic, and self-conscious instance of the problem-solving approach. To argue that deductive procedures permeate the entire universe of knowledge and subject all problem-solving to them would seem to me to move in the opposite direction – mythologizing daily problem-solving. Having pointed out the limits of method, Popper joyfully accepted what to others seemed as anarchy. Such anarchy, he thought, actually amounted to rational public debate that will lead to improvement. He joined the call for logical precision and methodological control with recognition that rational exchange cannot be governed by them. There was no cause for despair. When participants listened to each other and were eager to learn, they could achieve ever greater clarity of the issues at stake.

Rational control increases clarity. Arguments should conform to deductive reasoning as closely as is possible, so as to facilitate discussion and increase the efficacy and precision of criticism.[9] But wide space should be left open to proposals that rarely (and, in politics, almost never) reach deductive formulation because free exchange, at whatever level of precision, is potent. Recognizing this much, however, means a shift in focus from deductive reasoning to public debate. As wide areas of scientific and political discourse do not abide by deductive reasoning, yet are valuable to problem-solving, we need to find new ways of speaking about them and to explore non-methodological means of dis-

tinguishing better from worse arguments. Moreover, it is incumbent on us to think how we construct the public sphere so that it facilitates critical debate, the competition of proposals (and values), however imprecisely formulated. In this endeavor, we can derive little instruction from logic or method. We must call on history, sociology, psychology, and other social fields and derive whatever knowledge we can from them on the character of public debate and the setting most conducive to it.

I recounted the progressive marginalization of deductive method in Popper's philosophy by tracing the changing character of scientific and public debate in his philosophy and delimiting the role that method could be expected to play at each stage. My story has remained primarily one of internal theoretical development. I made few references to broader historical contexts to illuminate the transformation in Popper's views. This moderate historicization was sufficient, however, to address the concerns of both the scientists and science critics – Stadler's protagonists. The marginalization of method in the philosophy of its most formidable twentieth-century exponent ought to fill scientists with humility about the prospects of timeless rules in the pursuit of knowledge. But there is no cause for science studies relativism – critical debate safeguards objectivity (intersubjectivity) and rationality, and deductive method remains an invaluable helper, a facilitator, of public discussion, encouraging clarity, precision, simplicity. Stadler's goal is reached: History helps shape a universe of fallibilist knowledge by calling attention to the limits of method and, yet, vindicating the possibility of rational discourse in science and politics alike.

But can this vision of rational discourse survive a more radical historicization than the one endeavored here, one which grounds theories and methodological practices in particular contexts or denies them a role in scientific accomplishments? I think it can. The historically bound character of logic, method, and other scientific conventions has no bearing on rationality so long as critical debate is deemed potential arbitrator. Such debate remains a regulative ideal. Science studies will have, however, the last laugh, if they show that criticism has rarely had the beneficial effects that Critical Rationalists bestow on it. This depends partially on Critical Rationalists. If they remain wedded to logic and method and neglect the fields where science studies must be answered, the game is lost. They must show how, historically, critical dialogue propelled scientific advances and draw reform visions, imagining alternative scientific communities where democratic critical exchanges play greater role than they do in contemporary science.

NOTES

1. Rudolf Carnap, *Logische Syntax der Sprache* [1934], 2nd ed. (Vienna: Springer, 1968), p. VII, as quoted in Stadler.
2. (Tübingen: Mohr, 1979).

3. *Logik der Forschung* (Vienna: Springer, 1935), p. 61; *The Logic of Scientific Discovery,* trans. Popper (London: Hutchinson, 1959), p. 104: "If someday it should no longer be possible for scientific observers to reach agreement about basic statements, this would amount to the failure of language as a means of universal communication ... a new 'Babel of Tongues.' ... The soaring edifice of science would soon lie in ruins."
4. *The Open Society,* 2 vols. (London: Routledge, 1945): I:109-111, II:204-6, 224-6.
5. Recent studies point in this direction: Ian Jarvie, *The Republic of Science* (Amsterdam: Rodopi, 2001); Jeremy Shearmur, *The Political Thought of Karl Popper* (London: Routledge, 1996).
6. I am indebted to an email exchange with David Miller (July-August 2002) and to a conversation with Karl Milford during the ESF workshop (July 8, 2002). (Neither Miller nor Milford shares, however, the argument concerning background knowledge.)
7. Joseph Agassi writes (email of April 2, 2003): "This is true even of contemporary mathematics."
8. David Miller, "Reasoning: Control, not Command", http://www2.warwick.ac.uk/fac/soc/philosophy/staff/miller/dm_control.pdf, a response delivered at the seminar "Induction: Fact or Fantasy?" held at King's College London on January 9, 2002. See also his "Being an Absolute Skeptic," *Science* 284:5420 (4 June 1999): 1625f.
9. Joseph Agassi writes (email of April 2, 2003): "You are right, but you should credit it to Popper."

Department of History
Duke University
Durham, NC 27708-0719
U.S.A.
mhacohen@duke.edu

THEO A.F. KUIPERS

INFERENCE TO THE BEST THEORY, RATHER THAN INFERENCE TO THE BEST EXPLANATION – KINDS OF ABDUCTION AND INDUCTION [1]

1. INTRODUCTION

An interesting consequence of the structuralist theory of truth approximation, as developed in my *From Instrumentalism to Constructive Realism* (2000), henceforth ICR, concerns so-called 'inference to the best explanation' (IBE). It will be argued that this popular rule among scientific realists can better be replaced by, various kinds of, 'inference to the best theory' (IBT).

1.1. Background

I will start this paper with a survey of the main conclusions of ICR. It will provide the background of the present paper, which is partly even drawn from ICR, to be indicated, and partly supplementary, in particular the theme of 'abductive justification'. The survey of ICR itself is an adaptation of ICR-section 13.1. The conclusions that are the most relevant for this paper are preceded by a *. As a rule, these conclusions will be elaborated to some extent in the rest of the paper.

*The *first* conclusion, from Part I and II of ICR, is that a useful distinction can be made between two kinds of applying the Hypothetico-Deductive (HD-) method, viz., testing of hypotheses, in order to try to establish their truth-value, leading to falsification or confirmation, and evaluation of theories, in the first place, in order to establish their observational merits and failures.

Second, according to Part I, the notion of confirmation is primarily related to hypothesis testing, and can be given a direct qualitative explication, presented in Ch. 2, at least as far as deductive confirmation is concerned. Moreover, a quantitative, Bayesian, explication, given in Ch. 3, leads to a clear division of deductive and non-deductive confirmation. This quantitative explication indirectly generates a qualitative explication of confirmation in general and of non-deductive confirmation in particular. For the acceptance of observational hypotheses as true, fallible, inductive jumps have to be made in the sense of inductive generalizations. In Ch. 4 it is shown that inductive confirmation and

F. Stadler (ed.), Induction and Deduction in the Sciences, 25–51.
© 2004 *Kluwer Academic Publishers.*

inductive logic pertain to a particular kind of probability functions governing quantitative, deductive and non-deductive, confirmation.

**Third*, the evaluation of theories by the HD-method, studied in Part II, concerns, in the first place, the separate evaluation in terms of general successes and counter-examples, but immediately suggests a comparative sequel, leading to the establishment of empirical progress by applying the Rule of Success. The resulting evaluation methodology can be seen as an explication of the instrumentalist methodology, according to which counter-examples play a crucial role, but not the dramatic one that is associated with the falsificationist methodology. This explains and even justifies non-falsificationist, i.e., dogmatic, behavior of scientists. However, in contrast to pseudoscientists, they nevertheless aim at the improvement of their theories, which is frequently possible within a set of dogmas.

**Fourth*, empirical progress cannot be seen only as an aim in itself; it can also be seen as a means to truth approximation, as argued from Ch. 7 onwards. Here, 'the truth' is conceived as the nomic truth, provided by the so-called Nomic Postulate, i.e., the strongest true theory about what is nomically possible for a given domain of reality assuming a certain vocabulary. Using the structuralist approach to theories, the Nomic Postulate assures the existence of a subset of nomic possibilities within the set of conceptual possibilities generated by a vocabulary. Within this framework, a *prima facie* plausible definition of the idea that one theory is more truthlike than another can be given. The latter can be seen as a hypothesis, the truth approximation hypothesis, which can be tested due to its prediction of empirical progress, which makes empirical progress functional for truth approximation.

Fifth, the idea of more truthlikeness and the aim of truth approximation can be shown, in Ch. 8, to be in accordance with scientific common sense about theory improvement, viz., more true consequences and more correct models. Moreover, they suggest intralevel explications of some persisting intuitions among philosophers, viz., the correspondence theory of truth and intuitions governing dialectical reasoning.

**Sixth*, the (shifting) distinction between observational and theoretical terms not only naturally leads, together with the Nomic Postulate, in Ch. 9, to the distinction between theoretical and observational truthlikeness, but also to definitions of reference, referential claims of theories, and referential truthlikeness. Whereas observational truthlikeness implies empirical progress, theoretical truthlikeness only implies observational truthlikeness as far as true consequences are concerned. This leads to a further relativization of the role of counter-examples. Whereas the loss of a true observational consequence is a genuine drawback of a new theory, an extra counter-example of a new theory may well be an accidental (observational) merit of the old one. Moreover, since there are no deductive relations between referential truthlikeness and the other comparative notions, other kinds of (theoretical and experimental) arguments may also have to be taken into account for the assessment of referential claims.

Seventh, although the analysis sketched above may come close to scientific common sense, it does not yet mean that the crucial notions are directly applicable to real-life scientific examples. However, in Ch. 10 it is shown to be perfectly possible to take into account that the improvement of theories is usually not a matter of replacing mistaken models by correct ones, but of replacing mistaken models by less mistaken ones, e.g., when using the method of idealization and concretization, as for instance Van der Waals did in two concretization steps.

Eighth, Ch. 11 demonstrates that the succession of the (old) quantum theories of the atom of Rutherford, Bohr and Sommerfeld provides an example of potential truth approximation, more specifically, by specialization followed by concretization. A succession of theories about the capital structure of firms, the theory of Kraus and Litzenberger, followed by that of Modigliani and Miller, illustrates a different kind of truth approximation, but one which is, formally speaking, again by double concretization.

Ninth, and finally, speaking of distance from the truth and truth approximation easily suggests quantitative explications, as presented in Ch. 12. Although this is formally possible, assuming that there is some kind of distance function between the conceptual possibilities, it only seems to make sense in contexts where such a distance function is plausible.

Many other conclusions followed from the above ones, but I will not present them here, but confine myself to one, *tenth*, overall epistemological conclusion, which I will first introduce. The nine already mentioned conclusions roughly correspond to the division of ICR into four parts. More specifically, the first conclusion combines those of Part I about Confirmation and those of Part II about Empirical Progress, that is, the second and the third conclusion, respectively. The fourth to the sixth conclusion pertain to Part III about Basic Truthlikeness, and the seventh to the ninth conclusion pertain to Part IV about Refined Truthlikeness. However, ICR opens and closes with general epistemological chapters.

Ch. 1 introduces the main epistemological positions, instrumentalism[2], constructive empiricism, referential realism, constructive (theory) realism and essentialist (theory) realism. Ch. 13 provides a further articulation of constructive realism. The reason is that at the end of Part III the following conclusion was already drawn provisionally and was further strengthened in Part IV:

**Tenth*, the instrumentalist methodology provides good reasons for the transition of epistemological positions from instrumentalism, via constructive empiricism and referential realism, to constructive realism. Here, the intermediate step from constructive empiricism to referential realism turned out to be the hardest one, whereas the step from constructive to essentialist realism had to be rejected.

In the rest of Ch. 13 I have drawn the main lines of the resulting favorite epistemological position of *constructive realism*. It is a conceptually relative, hence non-essentialist, nomic truth approximation version of theory realism, accounting for objective truths. They can be approached by an intersubjective method, viz., the evaluation methodology for theories, in which the role of

(truth-)testing of hypotheses primarily concerns testing test implications of theories as well as testing comparative success and truth approximation hypotheses between theories.

The term 'constructive realism' has earlier been used by Giere (1985), and my conception of it is rather similar, except that I include in it, of course, truth approximation, whereas Giere still focuses on the true/false dichotomy, but he fully recognizes the nomic aim of theorizing. With respect to truth approximation, my position is rather similar to that of Niiniluoto (1987). The main difference between my and his position is, besides my primarily qualitative versus his straightforward quantitative approach, my emphasis on the nomic aim of theorizing. In sum, my constructive realism reflects the combination of their deviating strengths by emphasizing *nomic truth approximation* as opposed to the *actual truth-value* of theories.

1.2. Terminology

Before giving an outline of the paper, I will first delineate some crucial terms and types of conclusions. The terms 'induction' and 'inductive argument' will be used in the broad sense (Thagard, 1988, Salmon et al., 1992/1999) of an argument that is ampliative.[3] Following Salmon (p. 11), it is an argument of which the conclusion goes beyond the content of its premises. Hence, it is not necessarily truth preserving, it is not erosion-proof, that is, new premises may (completely) undermine an inductive argument[4], even a strong one, and inductive arguments come in different (informal or even formal) degrees of strength.

The terms 'abduction' and 'abductive argument' will be used to refer to the special case that an inductive argument can be presented as a straightforward case of abduction in the minimal sense of Peirce captured by the famous quote [Peirce, CP, 5.189]:

The surprising fact C is observed.
But if A were true, C would be a matter of course.
Hence, there is reason to suspect that A is true.

This quote represents a minimal kind of abduction, even if we assume, as Peirce seems to have intended, that the second premise is to be interpreted as stating that A would explain C in a deductive sense.[5] Hence, abduction is a provisional, explanatory kind of induction, that is, it is an induction of a deductive-explanatory nature of which the conclusion has to be further tested. Frequently a couple of additional conditions are suggested, also by Peirce. To begin with, the hypothesis A should be (further) empirically testable, which I will assume throughout. Other possible extra conditions are: 1) the generation of hypothesis A is also part of abduction, 2) A has to be 'economic' in some absolute and/or comparative sense, and 3) the conclusion can be made stronger, e.g. "Hence, we may assume that A is true". Accordingly, the suggested modest sense does

neither imply a discovery claim, nor an economic claim regarding A, and suggests the need of further testing A.[6]

Inductive arguments and, more generally, rules of inductive inference, that is, general patterns of inductive arguments, will be said to have an 'abductive justification' if a (re-)formulation of the argument as an abductive argument is possible. Otherwise I will speak of merely an 'inductive justification'.

Regarding types of conclusions of inductive arguments, it is important to note that in the context of theories and truth approximation there are at least four interesting conclusions. It may be concluded that (there is reason to suspect that) a theory is

1. true
2. the closest to the truth of the available alternatives
3. true and the closest to the truth of the available alternatives
4. the truth

Here, 'the truth' is conceived as the strongest true theory about a given domain in a given vocabulary. More specifically, 'the truth' (and 'true') may concern some actual possibility (world), 'the actual truth', or all 'nomic possibilities' in the domain, 'the nomic truth'. The latter target will here be assumed throughout, except when otherwise stated.

Of course, conclusion 4 entails conclusion 3, which entails, by definition, conclusions 1 and 2. Moreover, conclusions 1 and 2 are logically independent. Note also that conclusion 4 can be given, like conclusion 2, a comparative, even superlative, formulation: the closest to the truth of all possible theories. Conclusion 1, in contrast, is merely of a positive degree, and conclusion 3 is of a mixed nature. Instead of referring to numbered conclusions, I will also speak of conclusions of 'type-1, -2, -3, -4'.

1.3. Outline of the paper

I will first introduce, in Section 2, the unstratified rule of 'inference to the best theory' (IBT), that is, IBT with the type-2 conclusion 'as the closest to the truth'. It turns out to have a straightforward abductive justification, in contrast to (the rule of) 'inference to the best explanation' (IBE), that is, IBE with the type-1 conclusion 'as true', with which it will be compared. In Section 3, observational and theoretical kinds of both rules will be defined, compared, and discussed in the light of the epistemological positions, with essentially similar results. Section 4 deals with the prospects for referential truth approximation and the IBT-like rule of 'inference to the referential claim of the best theory' (IrcBT), that is, IrcBT 'as the closest to the referential truth'. It will be compared with 'inference to the likeliest cause' (ILC), that is, ILC 'as true'. In Section 5, I will introduce and discuss weak and strong forms of three kinds of induction with type-1 and

type-4 conclusions, respectively: observational, theoretical, and referential induction, and compare them with corresponding versions of IBE/ILC in their type-1 reading as well as in their type-3 (= type-1 + type-2) reading. I will close with some general conclusions and surveys (Section 6) and a discussion (Section 7) of points raised by Adam Grobler and other participants of the ESF-conference.

As far as possible, I will formulate the main points and arguments independent of the specific formal analysis of nomic truth approximation in ICR. Ideally, they apply, perhaps in some modified way, to any comparative theory of *nomic truth* approximation, and perhaps even to any comparative theory of *actual truth* approximation as well.

2. UNSTRATIFIED INFERENCE TO THE BEST THEORY

2.1. Preliminaries

In ICR-section 7.5.3., the first plea for reshaping 'Inference to the best explanation' (IBE) as 'Inference to the best theory' (IBT) is given. It is still unstratified in the sense that there is not yet made a distinction between observational and theoretical terms. It presupposes explications of the notions of 'being at least as (more) successful' and 'being as close (closer) to the truth'. The first can be explicated in at least three, strongly related, ways in terms of successes and problems of separate theories resulting from straightforward applications of the HD-method. In the present context I prefer more specifically:

> Theory Y is (at time t) *at least as (more) successful as (than)* theory X iff (at t)
> – all individual problems of Y are individual problems of X
> – all general successes of X are general successes of Y
> (– X has at least one extra individual problem or Y has at least one extra general success)

From this perspective, the crucial methodological steps are the following. Suppose that Y is, at t, more successful than X. This suggests the *Comparative Success Hypothesis*:

> CSH: Y (is and) remains more successful than X

CSH has of course to be further tested, now by the comparative application of the HD-method. At last we may come to the conclusion that it has been sufficiently tested, which provides the condition for applying the *Rule of Success*:

RS: If CSH has been sufficiently confirmed to be accepted as true, i.e., if *Y* has so far proven to be more successful than *X*, then eliminate *X*, in favor of *Y*, for the time being.

By 'rRS' I will indicate RS restricted to not yet falsified theories. The phrase "if *Y* has so far proven to be more successful than *X*" may be made more precise in Popperian style by adding: "despite severe attempts to falsify CSH".[7]

Fitting to the explication of 'more successful' is the following definition of 'closer to the truth':

Theory *Y* is *at least as close (closer) to the nomic truth T as (than)* theory *X* iff
– all missing models of *Y* are (missing) models of *X*
– all true consequences of *X* are (true) consequences of *Y*
(– *X* has at least one extra missing model or *Y* has at least one extra true consequence)

As already alluded to, 'the (nomic) truth T' is here interpreted as the unknown but fixed set of *nomic* possibilities, which is posited by the '*Nomic Postulate*', given a fixed domain and a fixed vocabulary. The vocabulary is designed to characterize the domain and gives rise to the set of *conceptual possibilities*. In this stage I assume that the vocabulary is observational, that is, all non-logical terms can be applied in an intersubjective way. Of course, observational terms may well be, and usually are, laden with background theories.

Like CSH, the *truth approximation (TA-)hypothesis*, *Y* is at least as similar to the truth *T* as *X*, is a decent comparative hypothesis. Assuming that the data are correct, it can even explain that *Y* is at least as successful as *X*, for the following theorem is easy to prove:

Success Theorem:
If theory *Y* is at least as close to the truth *T* as *X* and if the data are correct then *Y* (always) remains at least as successful as *X*

By this theorem, HD-testing of the TA-hypothesis immediately leads to (further) testing of CSH.

Finally, we have a theorem in the other direction:

Forward Theorem:
If CSH (which speaks of *remaining* more successful) is true, this implies the TA-hypothesis.

Given an observational vocabulary, this theorem arises from the fact that, if *Y* is not closer to the truth than *X*, (further) testing of CSH will sooner or later lead to

an extra counter-example of Y or to an extra success of X. The Forward Theorem leads to the claim that RS, and hence the HD-method, are functional for approaching the observational truth.

By now we have all the ingredients for ICR-section 7.5.3., which I will just repeat here, only with some marginal adaptations.

2.2. *Inference to the best theory, rather than the best explanation*

As suggested, an interesting consequence of the truth approximation analysis concerns so-called 'inference to the best explanation'. I will argue that it has to be replaced by 'inference to the best theory'. Let me first define the two crucial notions: 'the best theory' is the best, i.e., the most successful, theory among the available theories, assuming that there is one, whereas 'the best explanation' is the best theory which has not yet been falsified. Adapted to the present context, the standard idea of 'inference to the best explanation'[8] amounts to the following:

> *Inference to the best explanation (IBE)*:
> If an explanation has so far proven to be the best one among the available theories, then (choose it, i.e., apply rRS, and) conclude, for the time being, that it is true (i.e., that it is true as hypothesis)

In our setting, 'true as hypothesis' is plausibly interpreted as 'true with respect to all nomic possibilities', which amounts in the present context to the claim that the truth, that is, the set of nomic possibilities, is a subset of that explanation. IBE is problematic for at least three reasons. First, it does not say anything about falsified theories. It is restricted to the case that the best theory has not yet been falsified, hence counter-examples are not taken into account. Second, it has an asymmetric character in the sense that it couples a non-comparative (type-1) conclusion, being true, to a comparative premise, being the best unfalsified theory. Third, and finally, it is difficult to see how it could be justified in terms of general considerations concerning the true/false-distinction.

Consider the plausible alternative, suggested by RS and the Forward Theorem:

> *Inference to the best theory (IBT)*:
> If a theory has so far proven to be the best one among the available theories, then (choose it, i.e., apply RS, and) conclude, for the time being, that it is the closest to the nomic truth T of the available theories

IBT does not have the three shortcomings of IBE. It applies to unfalsified as well as to falsified theories. It couples a comparative (type-2) conclusion, being the closest to the truth, to a comparative premise, being the best theory. Last but not

least, it is directly justifiable in terms of truth approximation, viz., by the Forward Theorem. To be precise, the comparative hypothesis that the best theory remains the best, a hypothesis that apparently seems to be true, implies that the best theory is the closest to the truth. Hence, IBT can be seen as a, for good reasons, severely corrected version of IBE.

2.3. IBT as a kind of abduction

We might call the previous justification of IBT an 'inductive justification', because it is essentially non-deductively arguing in favor of the truth of the (proper) conclusion. Now I will turn to an (explanatory or) 'abductive justification', as defined in the first section. Recall that Peirce's basic scheme for abduction reads as follows:

The surprising fact C is observed.
But if A were true, C would be a matter of course.
Hence, there is reason to suspect that A is true.

Let us see whether IBE and IBT can be viewed as cases of abduction in this sense. In the case of IBE we have to interpret C as the (surprising) observation that a certain theory has so far proven to be better than any of its available alternatives *and has not yet been falsified.*[9] Moreover, A has to be interpreted as this theory. However, it is immediately clear that the second premise of the abductive argument does not hold: although 'being true' entails 'not yet being falsified', it does not at all entail being, relative to its available alternatives, the best explanation of the available data, not even that it is at least as good an explanation as its alternatives.

Turning to IBT, we have to interpret C 'merely' as the observation that a certain theory has so far proven to be the best one among the available theories and A has to be interpreted as the hypothesis that this theory is closer to the truth than any of its alternatives. In this case, the second premise ("But if A were true, C would be a matter of course.") holds almost perfectly, mainly due to the Success Theorem. More specifically (ICR, p. 162):

the theorem enables the attachment of three conclusions to the fact that Y has so far proven to be more successful than X; conclusions which are independent of what exactly 'the nomic truth' is:

– first, it is still possible that Y is closer to the truth than X, a possibility which, when conceived as hypothesis, the TA-hypothesis, according to the Success Theorem, would explain the greater success,

– second, it is impossible that Y is further from the truth than X (an hence X closer to the truth than Y), for otherwise, so teaches the Success Theorem, Y could not be more successful,

– third, it is also possible that Y is neither closer nor further from the truth than X, in which case, however, another specific explanation has to be given for the fact that Y has so far proven to be more successful.

The first conclusion needs some qualification (which the second conclusion does not need). In the strict sense, the Success Theorem merely explains deductively that Y is at least as successful as X. However, although it does not entail that Y will have an extra success, it does entail this very possibility. Put in a different way, the surprising observation is not so much that Y is at a certain moment more successful than X (the assumed point of departure for testing CSH), but that it remains more successful and this is a matter of course, assuming the TA-hypothesis.[10] Moreover, as the third conclusion makes clear, without this hypothesis, this observation is far from a matter of course. On the contrary, it will be a hard job to explain it without this hypothesis.

In sum, unlike IBE, IBT is obviously a kind of abduction, even in the strong sense that the conclusion, that is, the TA-hypothesis, has been further tested, viz., via CSH. Of course, the inclination to draw the TA-conclusion may be seen as a good reason for still further testing of the TA-hypothesis, if possible also in other ways than via CSH.

Let me briefly further comment on the different types of conclusions of IBE and IBT. The typical and justifiable conclusion of IBT is a type-2 conclusion, that is, 'being closest to the truth'. We have also seen, suggested by IBE, that the type-1 conclusion, 'being true', cannot easily be seen as the conclusion of an abductive argument, assuming a comparative surprising observation. However, this does not exclude that such a conclusion can be 'sufficiently' supported by observations. But in that case, it is merely a kind of 'induction'. I will elaborate the various possibilities, after introducing the distinction between observational and theoretical terms. Then we will also see that the type-4 conclusion, 'being the truth', cannot only be seen as a kind of induction but even as an extreme special case of IBT.

3. STRATIFIED INFERENCE TO THE BEST THEORY

3.1. Stratification

In Ch. 9 of ICR, I have presented a refined version of the theory of truthlikeness and truth approximation by introducing a (relative) distinction between theoretical and observational terms. More specifically, assuming a domain of interest D, I first introduce an *observational (o-)level* and a *theoretical (t-)level* for D. Let Vt indicate the richest vocabulary of non-logico-mathematical terms that play a role in one of the theories involved, that is, all relevant observational and theoretical

terms which are in consideration. Let *Vo* indicate the subset of *Vt* of observation terms.

Let *Mp(Vo)* and *Mp(Vt)* indicate the set of conceptual possibilities on the o- and t-level, respectively. The *projection function* π is generally defined as the function from a set of conceptual possibilities *Mp(V)* of some vocabulary *V* onto that set *Mp(V')* of a sub-vocabulary *V'* for which holds that, for all *x* in *Mp(V)*, $\pi(x)$ is the result of dropping the extra components in *x* and the structure clauses using them. By πX I indicate the set of 'projections' of all members of *X*, and by $\pi^{-1}\pi X$ the set of 'reproductions' of these projections, of which *X* must be a subset, possibly a proper one, as is easy to check.

Applying the Nomic Postulate on both levels, these sets generate, in combination with *D*, unique, time-independent subsets of nomic possibilities on each level, called *the observational (nomic) truth To* and *the theoretical (nomic) truth Tt*, respectively. I will further assume that $\pi Tt=To$, the *Truth Projection (TP-) Postulate* (from the t-level to the o-level). That πTt is a subset of *To* is a semantic fact, for if a conceptual possibility on the t-level is nomically possible, it will remain nomically possible when one skips some components. For the motivation for the substantial side of the TP-Postulate, viz., *To* is a subset of πTt, see ICR p. 215-7.

It would be nice when, not only 'the truth' but also 'more truthlikeness', of e.g. *Y* relative to *X*, would be projected unconditionally. However, although it does hold to a large extent, we need a condition to really guarantee this projection. The condition amounts to the supposition that theory *X* does not have extra 'potential lucky hits', that is, observationally correct models of *X*, but not of *Y*, that have no theoretically correct models in *X-Y*. Formally, $(\pi X \cap \pi Tt)\text{-}\pi Y$ has to be empty. A general, *Y*-independent, hence stronger, way of excluding such potential lucky hits is assuming that *X* has no potential lucky hits at all or, in the terms of ICR, *X* is *relatively correct*, which formally amounts to the condition: $\pi(X \cap Tt) = \pi X \cap \pi Tt$. But here I prefer to use the weakest condition in the relevant theorems.

Projection Theorem:
if *Y* is at least as close to the theoretical truth *Tt* as *X*, and if *X* has, relative to *Y*, no extra potential lucky hits, then πY is at least as close to the observational truth *To* as πX

Together with the Success Theorem, that is, a generalized version, formally taking theoretical terms into account (see the next subsection for an adaptation of a similar adaptation of (observational) IBT), we now get:

Success Corollary:
if *Y* is at least as close to *Tt* as *X*, and *X* has no extra potential lucky hits, then *Y* is always at least as successful as *X*

The Forward Theorem remains of course as before, apart from formally taking theoretical terms into account. Now we have in addition:

Upward Theorem:
if πY is closer to To than πX then Y^* is closer to Tt than X^*, where
$X^* =_{df} X\text{-}\{x \in X\text{-}Y / \pi(x) \in \pi Y \cap To\}$ and $Y^* =_{df} X \cap Y$

Note that both X^* and Y^* are strengthenings of their corresponding departure theory. X^* omits all X's extra potential lucky hits, if any, and Y^* restricts Y to its intersection with X. Of course, the latter may well be a drawback relative to Y itself, but the fundamental point is that we can construe in this way a genuine example of theoretical truth approximation starting from a case of observational truth approximation. This leaves room for the possibility that Y is also an example of theoretical truth approximation relative to Y^*, but this we cannot be sure of.

The Upward Theorem leads, together with (a generalized version of) the Forward Theorem, to the claim that RS, and hence the HD-method, are functional for approaching the observational truth, which in its turn is functional for approaching the theoretical truth (see ICR, p. 214, for a detailed motivation).

3.2. Stratified inferences

Now I will extend the previous analysis of IBT and IBE to the more standard stratification by an observational and a theoretical level. I start with a generalized version of the unstratified IBT. I call it now the observational IBT (IBT-o). Whereas the unstratified version assumes that the whole vocabulary is observational, this is no longer necessary, by jumping, whenever relevant, to the so-called 'corresponding observational theory', which refers of course to the projection of a theory on the observational level.

Inference to the best theory as the closest to the observational truth:
IBT-o: If a theory has so far proven to be the best one among the available theories, then (choose it, i.e., apply RS, and) conclude, for the time being, that the corresponding observational theory is the closest to the (nomic) observational truth of the observational theories corresponding to the available theories.

Its inductive justification in terms of the (equally generalized version of the) Forward Theorem remains the same. Similarly for its abductive justification in terms of the Success Theorem. But now we have of course also a rule on the theoretical level:

Inference to the best theory as the closest to the theoretical truth:
IBT-t: If a theory has so far proven to be the best one among the available theories, then (choose it, i.e., apply RS, and) conclude, for the time being, that it is the closest to the (nomic) theoretical truth of the available theories.

For the inductive justification of this theoretical IBT we not only need to appeal to the Forward Theorem but also to the Upward Theorem, which leaves room for the possibility that the relevant truth approximation claim only applies to strengthened theories. However, I will not spell this out in further detail. For the abductive justification of IBT-t an appeal to both the Success Theorem and the Projection Theorem is required. The conditional character of the latter has an important relativizing consequence. For the explanation of 'being more successful' we need to assume not only 'being closer to the truth' but also that the worse theory has no extra 'potential lucky hits'.

Recall the severe corrections IBT contains in comparison to what is standardly called 'Inference to the best explanation (as true)' (IBE), where the best explanation is conceived as the best not yet falsified theory. Again one can distinguish an observational and a theoretical kind, where I restrict myself to the latter, for the former is again merely a generalized version of the unstratified one.

Inference to the best explanation as true (even on the theoretical level):
IBE-t: If an explanation has so far proven to be the best one among the available theories, then (choose it, i.e., apply rRS, and) conclude, for the time being, that it is true (i.e., it is entailed by the theoretical truth)

The corrections of IBT-t compared to IBE-t are the following: unlike IBE-t, IBT-t is not restricted to unfalsified theories, and IBT-t amounts to a comparative conclusion attached to a comparative premise, whereas IBE-t amounts to an absolute conclusion attached to a comparative premise. Finally, IBT-t is directly justifiable in terms of truth approximation, viz., inductively by the Forward Theorem and the Upward Theorem, and abductively by the Projection Theorem and the Success Theorem. Neither justification of IBE-t is easy to imagine.

3.3. Epistemological positions and IBT

Let us now investigate to what extent IBT may be acceptable for what I see as the four main epistemological positions, viz., instrumentalism, constructive empiricism, referential realism and theory-realism (of which I indicate in ICR a defensible, constructive version and a problematic essentialist version). In view of the Forward Theorem, IBT-o may well be acceptable for all positions as far as they are willing to take the notion of the nomic truth on the observational level seriously, which however does not seem to be the case for, for example, Van Fraassen (1980, 1989), who wants to avoid '(physical or nomic) possibility-talk'.

IBT-t will be acceptable for the theory-realist, assuming that he is willing to speak of true or false claims when there may be fictions involved, see below.

What about the referential realist? As suggested before, he, or she, may be willing to subscribe to IBT-o, but it is unlikely that somebody like Cartwright (1983, 1989) will subscribe to IBT-t with some enthusiasm. On the basis of my analysis, she might conclude that IBT-t is defensible and hence innocent, but she will see it nevertheless as unimportant, for she will think that there are good reasons that it will almost never have interesting applications. Her view seems to suggest that the fundamental laws of interest are not only false, but, at least when separately conceived, that they are far from the truth. Moreover, for the referential realist, the crucial question concerns, of course, whether there is something like 'inference to the most likely cause' (Cartwright, 1983, 1989), or in my ICR-terminology, and more generally, 'inference to the best (total) referential claim'. The background is the following one. In ICR I have introduced an intermediate level determined by the terms in *Vt* that refer, together denoted by *Vr*, where *Vo*-terms are supposed to refer anyway, but the reference of genuine theoretical terms is determined by whether they really matter in the transition from *To* to *Tt*, that is, whether they are merely fictions or not. This enables the definition of the referential claim of a theory, of 'the referential truth' as the strongest true referential claim, and of the notions of referential truthlikeness and truth approximation. This gives rise to a corresponding definition of the idea of inference to the best referential claim.

4. INFERENCE TO THE BEST REFERENTIAL CLAIM

4.1. Referential truth approximation

I will start with the most relevant passages from ICR (parts of Section 9.1.2, pp. 215-6, and Section 9.2.1, pp. 219-20, with only minor adaptations).

Between the observational and the theoretical level, I introduce the intermediate level of terms that really refer to something real concerning the domain, called the *referential (r-)level*. For term τ in *Vt*, let *Vt'* indicate *Vt*-{τ} and let π project *Mp(Vt)* onto *Mp(Vt')*. Then τ is said to refer according to *Tt*, or simply *really refers* or *Tt-refers*, if and only if *Tt*, being on formal grounds a subset of $\pi^{-1}\pi Tt$, is a *proper* subset of $\pi^{-1}\pi Tt$. That is, *Tt* is a proper subset of the reproduction of the projection of *Tt*. This clause, the *Tt-reference clause*, is not yet to be interpreted as a complete definition of reference. It is at most a partial definition, providing a reference criterion. It amounts to the claim that τ makes a difference or plays a substantial role in the theoretical truth *Tt* in the sense that *Tt* is constrained by τ in the sense defined. With this partial definition, I formally

operationalize the idea that there is in reality some (type of) item that plays the role τ is ascribed to by *Tt*.

The definition still needs some qualification. It seems adequate primarily for theoretical terms purporting to refer to attributes (properties, relations, and functions); attribute terms for short. However, for theoretical terms purporting to refer to classes of entities, used as domain-sets for attribute terms – entity terms for short – this definition does not seem to work. However, it is plausible to say that a (theoretical) entity term in *Vt* *Tt*-refers if and only if there is at least one (theoretical) attribute term that *Tt*-refers and that uses the entity term as (one of) its domain-set(s). The consequence is that an entity term does not *Tt*-refer if there are no attribute terms using it as domain-set. This is plausible, for in this case it is difficult to see how that theoretical domain could play a substantial role in *Tt*, which is precisely the reason for the detour via attribute terms. If this detour is not possible, the theoretical entities hang in the air as unconstrained entities, not distinguishable from genuine fictitious entities, at least not with the means provided by *D* and *Vt*.

In sum, my definition of reference amounts to a direct reference criterion for attribute terms and an indirect one for entity terms. Note that the very possibility to define reference of entity terms on the basis of the definition of reference of attribute terms is a good reason to extend the idea of 'entity realism' to 'referential realism'.

The given definition of reference is explicitly *Tt*-relative. We might abstract from this by the definition that a term τ refers in an absolute sense if there is at least one domain and if τ belongs to at least one vocabulary such that τ refers relative to the resulting theoretical truth. The resulting theory of reference is rather non-standard. It would be interesting to compare it with the standard theories. Although it evidently is not a causal theory, it is also not exactly a causal-descriptive theory in the sense of David Lewis.

For the moment I stick to the definition given above and define the referential vocabulary *Vr(Tt)* or simply *Vr*, called *the referential truth*, giving rise to the *referential (r-)level*, as the union of *Vo* and those members of *Vt* that *Tt*-refer. Hence, by definition, *Vo* is a subset of *Vr* and *Vr* is a subset of *Vt*. So I assume that it has already been established that the observation terms refer in some other context. They may or may not refer in the present context, i.e., they may or may not *Tt*-refer.

Let *Mp(Vr)* indicate the set of conceptual possibilities on the r-level. By applying the Nomic Postulate on this level, *Mp(Vr)* generates, in combination with *D*, a unique, time-independent subset of nomic possibilities, called the *substantial truth Tr*. Here I will neglect the topic of substantial truth approximation and its relation to observational and theoretical truth approximation (see the second part of ICR-section 9.1.2) and turn immediately to referential truthlikeness and truth approximation (ICR-section 9.2.1).

It is crucial to note that, whereas *Tt* determines which terms really or *Tt*-refer and which do not, that is, which are fictions, every theory implies in a similar

way one or more specific *Tt*-reference claims. For theory *X*, being a subset of *Mp(Vt)* with the claim "*X=Tt*", implicitly claims that τ in *Vt-Vo* *Tt*-refers if τ plays a substantial role in *X*, that is, if *X*, being on formal grounds a subset of $\pi^{-1}\pi X$ (where π again projects *Mp(Vt)* again onto *Mp(Vt-{*τ*})*), is a *proper* subset of $\pi^{-1}\pi X$.

As in the case of the definition of *Tt*-reference, I have again to qualify this definition because it only seems appropriate for theoretical attribute terms. For specific *Tt*-reference claims for theoretical entity terms we may again use the detour via specific *Tt*-reference claims for attribute terms: if τ in *Vt-Vo* is an entity term it *Tt*-refers according to *X* iff there is at least one attribute term, using that entity term as (one of) its domain-set(s), which is *Tt*-referring according to *X*.

Let us indicate the subset of *Vt* consisting of *Vo* united with the set of theoretical terms that *Tt*-refer according to *X* by *Vr(X)*. Hence, *Vo* is a subset of *Vr(X)*, which is a subset of *Vt*. However, we do not know how *Vr(X)* is situated with respect to *Vr*. I call the conjunctive claim of *X* that all τ in *Vo* refer, that all τ in *Vr(X)-Vo* *Tt*-refer and that all τ in *Vt-Vr(X)* do not *Tt*-refer *the referential claim of X*, and it will be indicated by RC-*X*. In this way, the referential truth becomes the referential claim of *Tt*, hence the strongest true referential claim about *Vt*.

For subsets X and Y of Mp(Vt) I define, of course, the comparative claim "theory Y is at least as close to the referential truth Vr as X", or "Y is referentially at least as close to the truth as X", if RC-Y is at least as close to the referential truth as RC-X in the following plausible sense:

for all τ in *Vr* if τ belongs to *Vr(X)* then it belongs to *Vr(Y)*
for all τ in *Vt-Vr* if τ belongs to *Vr(Y)* then it belongs to *Vr(X)*

It amounts to a case of propositional actual truthlikeness, as defined in Ch. 7 of ICR, ranging over all elementary propositions of the form: "τ refers" for all τ in *Vo* and "τ *Tt*-refers" for all τ in *Vt-Vo*.

4.2. Inference to the referential claim of the best theory

In the first part of this section I terminologically deviate somewhat from the corresponding part of ICR-section 9.3.1.

In the light of the analysis above, it is now plausible to formulate the rule:

Inference to the referential claim of the best theory as the closest to the referential truth (IrcBT):
If a theory has so far proven to be the best one among the available theories, then conclude, for the time being, that its referential claim is the closest to the referential truth

Hence, if *Y* is the best theory, in some *Vt*, then IrcBT suggests that we conclude that *Vr(Y)* is closer to *Vr* than *Vr(X)*, as defined above, for any available alternative theory *X* in *Vt*. My previous analysis suggests that IrcBT is plausible, though its conclusion is certainly, like the IBT-o/t conclusions, not compelling. However, it is without doubt more plausible than the straightforwardly generalized rule suggested by Cartwright's 'inference to the likeliest cause', where I can and will now leave room for already falsified theories:

> *Inference to the referential claim of the best theory (the likeliest cause) as true* (ILC):
> If a theory has so far proven to be the best one among the available theories, then conclude, for the time being, that its referential claim is true

There is only one exception to the superiority of IrcBT over ILC. They coincide when there is only one theoretical term at stake, which is *Tt*-referring according to the best theory. Otherwise, ICL is (much) stronger than IrcBT.

Let us turn to the more detailed inductive and abductive justification of IrcBT, starting with the second. Hence, the question is, what does the conclusion of IrcBT, that the referential claim of the best theory is the closest to the referential truth, explain? The problem is that this conclusion itself does not explain very much. For this reason, referential realists have a hard job to provide abductive arguments for their position. The problem is that we have to take the conclusion of IrcBT together with the conclusion of IBT-t itself into account, a move that is only available to the theory realist.

One might think that the conclusion of IBT-t, viz. the best theory is the closest to the theoretical truth, does entail the conclusion of IrcBT. A specific reason to expect this is the fact that the theoretical claim of a theory, e.g. "*X=Tt*", entails its referential claim "*Vr(X)=Vr(Tt)*". However, a proof for the suggested conjecture is not possible, for interesting reasons. Suppose that *Y* is closer to *Tt* than *X*. What we would like to prove amounts to the following: all τ in *Vr(Y)-Vr(X)* belong to *Vr* and all τ in *Vr(X)-Vr(Y)* do not belong to *Vr*. Let us suppose that τ in *Vr(Y)-Vr(X)* does not belong to *Vr*. Hence, *Y* wrongly claims that it does, whereas *X* rightly claims that it does not. However, *X*'s claim may be based on a wrong aspect of its theoretical claim, whereas *Y* may not yet be so good that its theoretical claim implies the reference claim for the right reasons. It is important to note that the suggested proof already fails in the case of just one theoretical term. Similar arguments can be given for the possibility that τ in *Vr(X)-Vr(Y)* belongs to *Vr*. In both cases we might say that *X* has lucky reference claims relative to *Y*, that is, where *Y* has unlucky ones. In sum, the conclusion of IBT-t does not entail the conclusion of IrcBT.

This does not mean, of course, that the first conclusion does not provide good arguments for supposing that the relevant theory is closer to the referential truth. For, as can be learned from the suggested proof attempt, the suggested entail-

ment is only violated when a theory bases its referential claim on wrong reasons in precisely the right direction, which would be rather exceptional because it is rather artificial. In this respect there is a similarity with the relation between truth approximation and novel facts (see ICR-section 7.5.1.): although truth approximation without novel facts is in principle possible, it is rather artificial and hence may be supposed to be exceptional.

Recall now that the Projection Theorem, projecting more theoretical truthlikeness on more observational truthlikeness, has some extra condition, notably the assumption that X has no extra lucky hits, relative to Y. Hence, the next question is whether the conclusion of IrcBT is of any help in filling this gap in the Projection Theorem. Since I do not yet have a proof I formulate the following conjecture:

Conjecture:
If Y is closer to Tt than X then the assumption that RC-Y is closer to RC-Tt than RC-X narrows down (or even excludes) the possibility for potential lucky hits of X relative to Y

Although so far proof attempts were failing, I also did not succeed in proving that it does not narrow down the possibilities for lucky hits. On the contrary, it seems very much that, as a rule, the gap is substantially reduced, though not necessarily completely. The idea is, of course, that potential lucky hits of X relative to Y may be due to lucky reference claims of X relative to Y, which are excluded by the closer to the referential truth assumption of Y relative to X. Hence, I provisionally conclude that the conclusion of IrcBT substantially strengthens the explanation of the superior success of the best theory on the basis of the Projection and the Success Theorem. That is, an abductive justification for IrcBT, in combination with that of IBT-t, seems possible. I will call such a justification an IBT(-t)-laden abductive justification.

As to the possible abductive justification of ILC, the situation is not as bad as that for IBT-t, for even if the best theory is already known to be false, its reference claim may well be true. However, an abductive justification remains impossible for it remains the case that ILC is essentially asymmetric, in such a way that its non-comparative conclusion is (much) more risky than the comparative one of IrcBT, for it is (much) stronger, except when only one term is in debate. If we assume the conclusion of IBE, that is, that the best theory, apparently not yet falsified, is true, then the conclusion of ILC does not have to follow, for the best theory may not be strong enough to entail the correct elementary reference claims. Only if there is just one term involved, and the best theory's reference claim is positive, which is plausible indeed, this has to be a correct claim if that theory is true. The final question is whether the surprising phenomenon, being the best theory, can be explained by assuming both the conclusion of IBE-t and that of ILC. That is, an IBE(-t)-laden abductive justification of ILC might be

possible. However, even two comparative premises cannot entail a non-comparative one.

In sum, not only IBT-o and IBT-t, but also IrcBT is likely to have an abductive justification, albeit an IBT-laden one, whereas not only IBE-o and IBE-t lack one, but also ILC lacks one, even an IBE-laden one.

Let us now turn to the inductive justification of IrcBT and ILC. In ICR-section 9.2.2. I have given a number of theoretical and experimental arguments that may support specific and general, separate and comparative, reference claims. As a matter of fact, the emphasis is on the prospects for (specific and general) referential induction, see the next section, that is, whether a certain theoretical term of a theory refers or all its theoretical terms refer rather than not refer according to the old, less successful, theory. This is the relevant question when the new, presumably, more successful theory postulates extra theoretical entities and properties, relative to the old theory, for in that case the 'closer to the referential truth' claim for this new theory implies the claim that the extra referential claims of the new theory are true. In other cases, however, the old theory will also have 'positive' reference claims, which are rejected by the new theory. For such cases, the arguments in ICR-9.2.2. have to be adapted. In general, I see the explicit theoretical and experimental arguments in that section for 'positive' referential induction, and the implicit ones for 'negative' referential induction, merely as a beginning. The only claim is, that my specific definition of reference in ICR is a plausible new point of departure.

However this may be, although one might think of other than abductive arguments in favor of IrcBT, it is difficult to imagine whatever argument in favor of ILC, simply due to its asymmetric character.

5. THREE TYPES OF INDUCTION

It is not only plausible to distinguish three kinds of induction, viz. observational, theoretical and referential, but also weak (type-1) and strong (type-4) forms of it.[11] Although I will not systematically speak of inductive jumps, this terminology would stress more strongly that inductions are always fallible, and that the standards of induction are debatable. Moreover, it would withdraw the suggestion that inductions are merely extrapolations, as in the case of straightforward inductive generalizations. But even such generalizations are essentially fallible leaps. Finally, in ICR-sections 3.3. and 4.3., I have argued that, although the (threshold) problem of the acceptance of hypotheses, as true or as the truth, is still a difficult one, it is relatively clear how acceptance is guided by the general and comparative principles of confirmation. Here I will concentrate on aspects that can now be formulated more clearly.

In the case of weak/strong *observational induction* it is concluded, for the time being, and for a certain domain and a certain observational vocabulary, on the basis of the available evidence, that a certain general hypothesis stated in that

vocabulary is true for all nomic possibilities (weak, type-1), or is the strongest true hypothesis about all nomic possibilities (strong, type-4), respectively. It is evident that observational induction requires the Nomic Postulate in order to bridge the gap between actual and nomic possibilities. To be sure, weak observational induction presupposes that the hypothesis has not been falsified, and strong observational induction presupposes *in addition* that (the strong claim of) the hypothesis is the best available theory. But these are just necessary conditions for observational induction, which requires, in addition, the much stronger and much more debatable assumption that enough attempts to falsify it, to improve it, respectively, have failed, in order to pass the relevant threshold of plausibility.

However, to call *weak* observational induction 'inference to the best explanation' (as true) would be misleading, since there is no comparison of hypotheses involved. This is not to say that observational induction cannot have abductive justification. On the contrary, this is perfectly possible. As a matter of fact, all 'deductively derivable' (more or less surprising) evidence, to begin with not being falsified, can be explained by its type-1 conclusion. Hence, cases of weak observational induction may well be called (cases of weak observational) abduction. Turning to *strong* observational induction, calling that 'inference to the best explanation' in the sense of 'inference to the best explanation as the (observational) truth' might be defensible as an extreme special case of IBT(-o), for strong observational induction amounts to 'inference to the best theory as the observational truth'. For this purpose, recall that, despite Note 11, a type-4 conclusion can be rephrased as a comparative, even superlative, conclusion: the closest to the truth of all possible theories. Moreover, 'being the truth' explains of course, even trivially, being not falsified and, by the Success Theorem, being the best theory of the available theories. Hence, cases of strong observational induction may well be called abduction.

Let us now turn to weak and strong *theoretical induction*. In this case, it is concluded, for the time being, and for a certain domain and a certain theoretical (cum observational) vocabulary, on the basis of the available evidence, that a certain general hypothesis stated in that vocabulary is true for all nomic possibilities (weak, type-1), or is the strongest true hypothesis about all nomic possibilities (strong, type-4), respectively. Again, it is evident that theoretical induction requires the Nomic Postulate. Moreover, weak theoretical induction presupposes again that the hypothesis has not been falsified, and strong theoretical induction presupposes, in addition, that (the strong claim of) the hypothesis is the best one of the available theories. As in the case of observational induction, these are now simply necessary conditions for theoretical induction, which require, in addition, the much stronger and much more debatable assumption that enough attempts to falsify it, to improve it, respectively, have failed, in order to pass the relevant threshold of plausibility. However, in ICR Ch. 2, I have argued that the guidelines provided by confirmation theory only fail in the context of theoretical

induction in the case of observationally equivalent *and* (initial and hence posterior) equally plausible theories.

Again, cases of weak and strong theoretical induction may be called cases of abduction, for similar reasons as in the case of observational induction, except that for the strong case the Projection Theorem has to be used as well for explaining to be the best theory. However, again, calling weak theoretical induction 'inference to the best explanation' (as true) would be misleading, since there is no comparison involved. But calling strong theoretical induction 'inference to the best explanation' in the sense of 'inference to the best explanation as the (theoretical) truth' might again be defensible as an extreme special case of IBT-t, for strong theoretical induction amounts to 'inference to the best theory as the theoretical truth'.

Let us finally turn to *referential induction*. Weak referential induction amounts to the conclusion, for the time being, that, under appropriate conditions, the conjunction of some specific referential claims of a theory, as far as they are positive, is true. In the case of strong referential induction it is concluded that the entire referential claim of the theory is true, that is, all specific positive claims as well as all specific negative ones. Recall that, although being closer to the theoretical truth did not guarantee being closer to the referential truth, the fact that the (strong) theoretical claim is true guarantees that the referential claim is true. Moreover, when the weak theoretical claim of a hypothesis is true, this guarantees that all positive specific referential claims implied by that weak claim are true. Accordingly, weak and strong theoretical induction imply, and hence are sufficient conditions for, weak and strong referential induction, respectively. Although it may be possible to formulate other (pragmatically) sufficient conditions for referential induction of a theoretical and experimental nature (see ICR-section 9.2.2.), I confine myself to referential induction as implied by theoretical induction. However, one should realize that one has to supplement theoretical induction explicitly with referential induction to take such other types of referential induction into account.

Weak and strong referential induction based on the corresponding theoretical induction obtain of course abductive justification via, and hence are laden with, such theoretical induction. However, calling weak referential induction 'inference to the likeliest cause' (as true) would be misleading, since there is no comparison involved. But calling strong referential induction 'inference to the likeliest cause' in the sense of 'inference to the referential claim of the best theory as the (referential) truth' might again be defensible, now as an extreme special case of IrcBT, for strong referential induction amounts to 'inference to the referential claim of the best theory as the referential truth'.

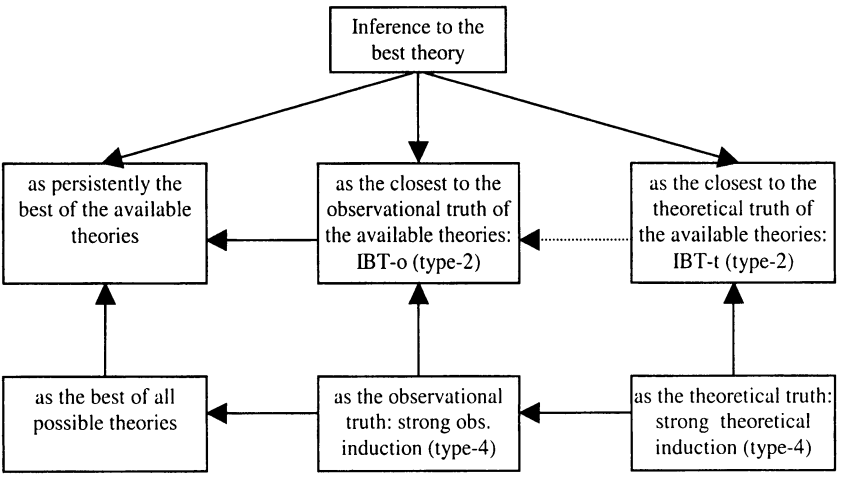
Finally, there is one (type of) way to explicate IBE and ICL that may be seen as the most benevolent and defensible. IBE might be interpreted as a combination of weak (observational/theoretical) induction and IBT(-o/t), and hence with a type-3 conclusion: inference to the best theory as (observationally/theoretically) true and as the closest to the (observational/theoretical) truth of the avail-

able theories. Similarly, ILC may be interpreted as a combination of referential induction and IrcBT, again with a type-3 conclusion: inference to the referential claim of the best theory as true and as the closest to the referential truth of the referential claims of the available theories. In all three cases, the supposition of not being falsified and the symmetric comparative character are built in, and hence both kinds abductive justification, that is, that for weak induction and that for IBT and IrcBT, respectively. Since abductive justifications of conclusions are essentially based on their explanatory merits, we get in this way interpretations of IBE and ILC that do justice to their names, in particular IBE. However, this does not imply that these combined inference rules add anything to their component rules. The latter, together with strong inductions, form the basic type of rules. To be precise, the three kinds of IBT and those of weak and strong induction seem to constitute the basic inductive rules, with abductive justification, for scientific realists.

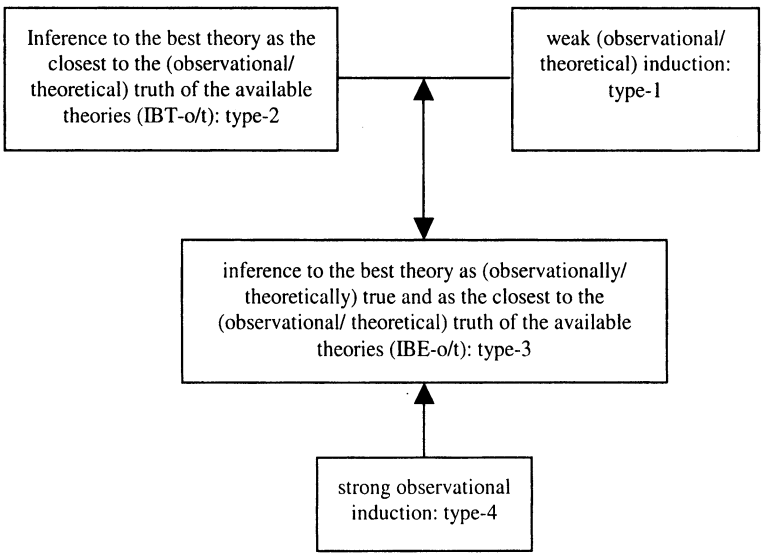
6. CONCLUDING REMARKS

The main conclusions of this paper are:

1) In the light of the structuralist theory of truth approximation, 'Inference to the best explanation' (IBE) and 'Inference to the likeliest cause' (ILC) should be replaced by kinds of 'Inference to the best theory' (IBT), that is, either by 'Inference to the best theory as the closest to the (observational, theoretical, or referential) truth' or by their extreme versions 'Inference to the best theory as the (observational, theoretical, or referential) truth', where the latter may also be called 'strong (observational, theoretical, or referential) induction'. In the diagram below I have left out the referential kinds. The arrows have plausible interpretations.



2) For the observational, theoretical, and referential (non-extreme and extreme) kinds not merely inductive justifications can be given but also abductive justifications based on the theory of truth approximation, albeit of decreasing strength.
3) IBE (observational and theoretical) and ILC lack abductive justification for several reasons, of which their asymmetric character is the common one.
4) In contrast, their non-comparative versions, here called 'weak (observational, theoretical, or referential) induction', have abductive justifications.
5) However, (1) and (4) can be combined to explications of IBE/ILC. Again I have left out the referential kinds in the diagram below and the arrows have plausible interpretations.



Finally, I like to conclude this paper with (an adapted version of the first paragraph of ICR-section 13.3) stressing the crucial role played by observational and theoretical induction in the construction and determination of terms and, hence, the long-term dynamics in science. Besides the formation of observation terms by straightforward explicit definition, observational induction may provide the necessary and sufficient conditions, e.g., in the form of existence and uniqueness requirements, for explicitly defining new observation terms. The quantitative notions of pressure and temperature are examples. Such new terms are unambiguously, hence intersubjectively, applicable, and they (may be supposed to) refer if they enable new observational inductions. Besides implying referential induction, theoretical induction may provide the necessary and sufficient conditions, e.g., in the form of existence and uniqueness requirements, for applying theoreti-

cal terms, that is, for identifying theoretical entities and for measuring theoretical attributes. E.g., the detection of electrons, and the measurement of their mass and charge, are based on such inductions. In this way, theoretical terms (may be supposed to) become referring and unambiguously, hence intersubjectively, applicable. In other words, theoretical terms can be essentially transformed into new observation terms by appropriate theoretical and/or referential induction. Of course, together with earlier or elsewhere accepted observation terms, they can be used for new cases of observational induction. Moreover, with other ones, they will play a crucial role in the separate and comparative evaluation of new theories introducing new theoretical terms, dealing with (partially) new domains, starting another round of the 'empirical cycle'.

7. DISCUSSION

In response to Adam Grobler's comments, I would like to briefly state the following points.

1) I agree that IBT should not be seen as an autonomous rule of inference characteristic for *the logic of science*, but as a heuristic rule characteristic for *the use of logic in science*. More specifically, I see it as an applied rule of inductive, hence non-monotonic, inference, building upon classical deductive logic, a given domain and vocabulary, and the model-theoretic or structuralist definition of truthlikeness.
2) I thank Grobler for making so explicit that applications of IBT, that is, inferences that the best theory is the closest to the truth, are only plausible, due to a couple of theorems, if it is assumed (not only that the data are correct, but also) that the background knowledge is true. Hence, questioning the background knowledge certainly is a way to avoid the truth approximation conclusion.
3) Grobler's suggestion, in response to my second complaint about IBE, to make its conclusion comparative by replacing it by "[the best theory] is the best candidate (among the alternatives) for the truth" is quite a step in the direction of my favorite IBT for two reasons. First, taking 'the truth' as the point of reference only makes sense if something like my definition of it is accepted, that is, the strongest nomically true theory. Second, assuming that we know that the best theory is false, if the data and the background knowledge are correct, what else could '[being] the best candidate ... for the truth' mean than 'being the closest to the truth'?
4) In contrast to Grobler's opinion, my definition of successes and failures includes explanatory successes and failures, at least when they are of a, strictly or approximately, derivational nature. However, it is not immediately clear whether explanatory derivational (general) successes are more functional for truth approximation than non-explanatory ones, assuming that the latter exist

at all. The last proviso is made, because 'flagpole' type of examples typically concern the explanation of individual facts and not general facts, such as general successes of theories. Be this as it may, as I concede in my reply there, Grobler and Andrzej Wiśniewski (to appear) are certainly right in suggesting that non-derivational types of explanatory successes, in particular successful 'explanations by (causal, functional, and intentional) specification', should be taken into account in the definition of the best theory. As I should have added in my reply, (correct) explanations by specification amount to specifying special kinds of true statements, hence it follows almost immediately that theories having more of such successes than others are likely to be closer to the (nomic or actual) truth.

5) I like to subscribe to Grobler's remark that 'the amount of incorrectness' may matter for the closeness to the truth. It is precisely this intuition that underlies the transition in ICR from the basic theory of truth approximation to the refined one of Ch. 10, which takes into account that one incorrect model made be more similar to a missing model (a counterexample) than another.

Let me also briefly discuss the comments made by other participants.

1) Asked for examples of straightforwardly more successful theories, I like to refer to a comparative table of seven electrodynamic theories in view of thirteen experimental results in (Panofsky and Phillips 1962[2], p. 282, reproduced in ICR, p. 118, and in Kuipers, 2001, p. 236).
2) Unlike in this case, in which the special theory of relativity dominates all other theories in success, it may of course be that there is no overall dominant theory, but that there are two or more 'locally dominant' theories, in which case the Rule of Success doesn't prescribe a choice between these theories.
3) Moreover, even in the case of an overall dominant theory, it may well be that this one is still very from the truth.
4) In general, if we succeed in improving a theory in the sense of banishing a counterexample in a non-ad hoc manner the Comparative Success Hypotheses will predict some novel facts, predictions that may or may not come true.
5) Although Mandelbrot is certainly right in that 'the truth' about the length of a certain coast does not exist, it does exist as soon as we use a minimal unit of length, that is, a vocabulary, as presupposed throughout.
6) Leaving room for inconsistent theories, as in paraconsistent logic, is a not merely a challenge for IBT, it requires an adaptation of the whole theory of truth approximation. This would be a very interesting research project.
7) As suggested in Notes 1 and 6, it is possible to take non-empirical success criteria into account in the definition of 'more successful' (Kuipers, 2002), and hence of IBT, of course, as far as this does not conflict with the empirical ones and as far as this is functional for truth approximation. Such criteria may not only depend on the discipline, and its stage of development, but

even on the style of scientific reasoning, as the historical evidence for an English and a French style seems to suggest. After all, there is more than one road to Rome. Be this as it may, regarding IBT there seems to be less need for including such non-empirical criteria than in the case of IBE, viz. in view of their rather different conclusions, for in the IBE-case additional arguments are needed for the non-comparative conclusion of being true.

NOTES

1. I would like to thank in particular Adam Grobler, Ladislav Kvasz, Joke Meheus, Ilkka Niiniluoto, Wolgang Spohn and Igor Douven for their stimulating comments on the draft version.
The idea of 'Inference to the best theory' is also discussed by Ilkka Niiniluoto (to appear) and by Igor Douven (to appear) in a 'volume in debate'. Niiniluoto's paper mainly deals with the unstratified kind and Douven's paper with the theoretical kind. In both Douven's paper and Kuipers (2002) arguments are given for including non-empirical, e.g. aesthetic, criteria in the comparative assessment of theories, with consequences for the precise formulation of IBT. See further my replies to Douven and Niiniluoto in the volume in debate referred to and my reply to Grobler and Wisniewski (to appear) in the companion volume.
2. Instrumentalism as an epistemological position should be distinguished from the instrumentalist or evaluation methodology. The latter is claimed to serve also other epistemological positions.
3. See Flach and Kakas (2000) for an interesting survey and discussion of the various uses of the terms 'induction' and 'abduction'.
4. Hence, an inductive argument is a non-monotonic argument with a default character: its conclusion is supposed to hold as long as there is no counterevidence.
5. However, weaker kinds, that is, non-deductive ones, are certainly possible, see e.g. Aliseda (1997, p. 27).
6. In the course of the paper, I will nevertheless frequently assume that (not only C, but also) A has already severely been tested. However, in this paper I will not deal with the generation of hypotheses, nor with economic factors. See Kuipers (1999) for an analysis of the main abductive generation and evaluation tasks that result from ICR. See Kuipers (2001, Section 11.1.4) for a sketch. See Kuipers (2002) for a way to take non-empirical factors, in particular aesthetic ones, into account.
7. Igor Douven has suggested an interesting weaker version of RS, by not prescribing to eliminate X in favor of Y when CSH has been sufficiently confirmed, but by prescribing something like: "spend more resources on Y than on X to the extent that CSH has been confirmed." This would legitimate some continued research attention to X.
8. Harman (1965) coined the term. See (Lipton 1991) for an extensive exposition and defence of IBE.
9. In contrast to Peirce's favorite examples, the surprising observation is here not an individual fact, but a kind of general fact, in the present IBE-directed interpretation *partly* of a comparative nature, in the IBT-directed interpretation, see below, *in full*.
10. In the IBE-directed interpretation the surprising observation would then be that the relevant theory 'remains more succcessful (than its alternatives) *and unfalsified*'. Again, the assumption that the theory (i.e., A) is true would make the second part of this observation a matter of course, but not the first part.
11. It might be argued that the three kinds of IBT discussed before are also kinds of induction, viz., comparative inductions, i.e., inductions with the comparative conclusion of being the closest to the relevant truth, in contrast to the separate or absolute kinds of induction discussed in this section, i.e., inductions with the 'absolute' conclusion of being true or of being the truth.

REFERENCES

Aliseda, A., (1997), *Seeking Explanations: Abduction in Logic, Philosophy of Science and Artificial Intelligence*. Dissertation Stanford. Amsterdam: ILLC Dissertations Series (1997-04).

Cartwright, N., (1983), *How the laws of physics lie*. Oxford: Clarendon Press.

Cartwright, N., (1989), *Natural capacities and their measurement*. Oxford: Clarendon Press.

Douven, I., (to appear), "Empirical equivalence, explanatory force, and the inference to the best theory", to appear in: R. Festa, A. Aliseda, and J. Peijnenburg (Eds.), *Confirmation, Empirical Progress, and Truth Approximation. Essays in Debate with Theo Kuipers*, Volume 1. Amsterdam: Rodopi.

Flach, P. and Kakas, A., (2000), "Abductive and inductive reasoning: background and issues", in: P. Flach and A. Kakas (Eds.), *Abduction and Induction*. Dordrecht: Kluwer AP, pp. 1-27.

Fraassen, B. van, (1980), *The scientific image*. Oxford: Clarendon Press.

Fraassen, B. van, (1989), *Laws and symmetry*. Oxford: Clarendon Press.

Giere, R. (1985), "Constructive realism", in: P. Churchland and C. Clifford (Eds.), *Images of science*. Chicago: The University of Chicago Press, pp. 75-98.

Grobler, A. and Wiśniewski, A., (to appear), "Explanation and Theory-Evaluation", to appear in: R. Festa, A. Aliseda, and J. Peijnenburg (Eds.), *Cognitive Structures in Scientific Inquiry. Essays in Debate with Theo Kuipers*, Volume 2. Amsterdam: Rodopi.

Harman, G., (1965), "The inference to the best explanation", *Philosophical Review*, 74, pp. 88-95.

Kuipers, T., (1999), "Abduction aiming at empirical progress or even truth approximation, leading to a challenge for computational modelling", *Foundations of Science*, 4.3, pp. 307-323.

Kuipers, T., (2000/ICR), *From Instrumentalism to Constructive Realism. On some relations between confirmation, empirical progress, and truth approximation*, Synthese Library, Vol. 287. Dordrecht: Kluwer AP.

Kuipers, T., (2001), *Structures in Science. Heuristic patterns based on cognitive structures*. An advanced textbook in neo-classical philosophy of science. *Synthese Library*, Volume 301. Dordrecht: Kluwer AP.

Kuipers, T., (2002), "Beauty, a road to the truth", *Synthese*, 131.3, pp. 291-328.

Lipton, P., (1991), *Inference to the best explanation*. London: Routledge.

Niiniluoto, I., (1987), *Truthlikeness*. Dordrecht: Reidel.

Niiniluoto, I., (to appear), "Abduction and truthlikeness", to appear in: R. Festa, A. Aliseda, and J. Peijnenburg (Eds.) *Confirmation, Empirical Progress, and Truth Approximation. Essays in Debate with Theo Kuipers*, Volume 1. Amsterdam: Rodopi.

Panofsky, W., and Phillips, M., (1962^2), *Classical Electricity and Magnetism*. London: Addison-Wesley.

Salmon et al., (1992, 1999), *Introduction to the Philosophy of Science*. Indianapolis/ Cambridge: Hackett Publishing Cy.

Thagard, P., (1988), *Computational philosophy of science*. Cambridge: MIT-press.

Faculty of Philosophy
University of Groningen
Aweg 30, 9718 CW Groningen
The Netherlands
T.A.F.Kuipers@philos.rug.nl

ADAM GROBLER

THE SIGNIFICANCE OF EXPLANATORY CONSIDERATIONS

Comments on Theo A.F. Kuipers's
"Inference to the Best Theory, rather than Inference to the Best Explanation"

To prevent possible misunderstanding, let me start from one crucial distinction between *the logic of science* on the one hand and *the use of logic in science* on the other. The very phrase "inference to the best theory" and the style of its further explication may suggest that it is the former that is at issue. In contrast, the title of the present workshop, "Induction and deduction in the sciences", suggests the latter to be its focus. And this I find to point to the more adequate way of thinking about the relation between logic and science. I do not believe in "the logic of science" as something that can possibly be read off science in a manner analogous to that of reading logic off mathematics: the classical logic off the classical mathematics or the intuitionistic logic off the intuitionistic mathematics.

There are at least two reasons for that, and they do not necessarily depend on the ampliative character of scientific reasoning. First, a mathematical theory starts from axioms, while in science there are no unquestionable premises of inferences. Consequently, there is nothing in science to which we may just apply the rules, from scratch as it were, and eventually draw conclusions. Thus, the conclusions of scientific inferences are always conditional on the premises which themselves are to be evaluated by scientific reasoning. Second, inferences are always relative to a certain vocabulary, what Kuipers explicitly admits. This is not a trouble as far as mathematics is concerned, for different vocabularies can be thought of as describing different, independent mathematical realities. But this is a serious problem in the philosophy of science, for the question of which vocabulary is more adequate as a tool of representing a particular domain is again to be decided by scientific reasoning.

How these two problems trouble the project of reconstructing "the logic of science" can be illustrated with the difficulties of Bayesianism. First, how a piece of evidence affects the probability of the hypothesis under test depends on background knowledge. Should background knowledge change, the preferences based on the probability estimates of rival hypotheses may change as well. The question is how a possible revision of background knowledge can be justified in Bayesian terms. Second, the engine of the probability calculus works within a particular probability space, i.e. within a particular vocabulary. The question of a

F. Stadler (ed.), Induction and Deduction in the Sciences, 53–56.
© 2004 *Kluwer Academic Publishers.*

conceptual change – of "a scientific revolution" – is again intractable in Bayesian terms.

Similar problems plague IBT as far as it is conceived as a rule of inference of the logic of science. First, IBT is formulated in terms of hypothetico-deductive successes and failures. What is a success or failure (or a neutral event) depends on background knowledge. Lakatos, i.a., gave us beautiful examples of how the failures of a scientific research program can be transformed into its striking successes. These can be thought of as examples of rejecting one and accepting another, incompatible with the former, IBT-conclusion, on the grounds that seem to be quite independent from IBT. Second, IBT-conclusions are relative to a certain vocabulary and hence IBT is not applicable to the question of evaluating the adequacy of competing vocabularies or conceptual schemes.

For these reasons IBT, just as the Bayesian conditionalization rule, and possibly any other proposal, is not a good candidate for a rule of inference of the logic of science. Rather, it is a heuristic rule that says that a hypothesis is to be tentatively accepted as the closest to the truth, in the absence of a better alternative relative to a certain background knowledge and vocabulary. As such, IBT is a product of a particular use of classical logic in science, namely its application to defining (background-knowledge relative) successes and failures of a theory or hypothesis. If this is correct, Kuipers' careful distinctions between four types of inferences, to truth, closeness to the truth, truth and closeness to the truth and the truth, seem to me a little misleading. They may wrongly suggest that it is a logic that is at stake, rather than a use of ordinary logic, a use limited to a particular background knowledge and vocabulary.

In the face of these limitations of ITB, the question arises whether it is possible to have other heuristic rules that would be helpful in deciding about a possible revision of background knowledge or vocabulary. In this connection, I am inclined to suggest that we reconsider IBE, contrary to Kuipers' claim about the superiority of the former over the latter. Indeed, in Grobler and Wiśniewski (forthcoming), we argued that Kuipers' own account of explanation can perfectly provide us with a quite important tool of theory-evaluation in terms of relative explanatory power, a tool supplementary with regard to the hypothetico-deductive method. Why then he is reluctant towards IBE? First, IBE is said to be silent about falsified theories. This feature does not prevent us from using IBE as a heuristic rule. Suppose that we arrive at a theory, otherwise falsified, such that it would be the best explanation in the domain if only it was not falsified. If we take explanatory considerations seriously enough, we may in the circumstances strive at a suitable revision of background knowledge in order to transform the theory's failure into its success, rather than to reject the theory and keep background knowledge intact.

The Kuipers' second complaint about IBE is that it leads to a noncomparative conclusion (truth) from a comparative premise (best explanation). But, similarly to my earlier suggestions concerning IBT, we may take the word "inference" not to imply that IBE is a rule of inference of a logic. Rather, IBE is to serve as a

heuristic rule. If this is so, the conclusion drawn about the best explanatory theory in a domain is to be read as "... is the best candidate (among the alternatives) for the truth" rather than "... is true" *simpliciter*. In this reading, the conclusion is precisely as comparative as the premise.

The Kuipers' last objection towards IBE is that it is not epistemologically justified with regard to truth. In my view, it again depends on its reading. If we take as "the best explanation" the one provided by a theory of the most inclusive class of its explanatory applications (for an early exposition of the concept see my 2000a, an enhanced version is contained in 2000b), then the best explanatory theory points to more interconnections between the phenomena than its rivals. Consequently, the best explanatory theory is the most informative one in the domain. Hence, if the best explanatory theory is true then it gives the most interesting truth in the domain and if it is false then it is more vulnerable to test and thereby its adopting is truth-conducive in the sense of the hypothetico-deductive method.

IBE gives us a more powerful tool than a purely instrumentalist methodology of IBT. First, explanatory successes and failures are always more impressive than instrumental ones. Next, even if explanatory considerations can be accommodated by the definition of the best theory, as Kuipers suggested in his reply to Grobler & Wiśniewski, IBE transcends IBT in that that it can be used as a heuristic rule for suggesting background knowledge revisions. One additional point in favor of IBE can be made in connection with defining IBT in terms of inclusion between the classes of correct and incorrect models of rival theories. As far as the closeness to the truth is concerned, the amount of incorrectness may matter. That is, even an incorrect model may count as a virtue of a theory rather than as its vice, provided that the model in question is explanatory useful. Such a criterion of the approximate correctness of a model is much more plausible than a possible criterion in terms of the degree of numerical accuracy of prediction, for in the latter case it is difficult to avoid employing conventional decisions.

The question of conceptual change, that is, of evaluating rival vocabularies for a given domain, can be handled in a way similar to that of a possible revision of background knowledge (that is, evaluating rival bodies of background knowledge). Here explanatory considerations can also play a role that IBT alone possibly cannot. On the other hand, IBE can here be fruitfully combined with Kuipers' idea of approximating the referential truth. Namely, the referential claim of a theory can be defined in terms of whether the presence of a term matters not only with regard to the theory's observational successes and failures but also with regard to its explanatory successes and failures.

To conclude, inference rules of the sort of IBT or IBE are to be conceived not as something analogous to the inference rules of logic. Instead, they are heuristic devices of theory choice that employ some techniques of logic as well as of set theory (inclusion relation between the sets of correct, approximately correct or incorrect models of rival theories, or of explanatory applications of rival theo-

ries) and possibly some other mathematical tools (probability measures, metrics, etc.). Including explanatory considerations seems to me indispensable, however.

REFERENCES

Grobler, Adam 2000a. 'Explanation and epistemic virtue' in Adam Jonkisz, Leon Koj (eds.) *On Comparing and Evaluating Scientific Theories, Poznań Studies in the Philosophy of the Sciences and the Humanities,* Vol. 72, pp. 61-88.

Grobler, Adam 2000b. *Prawda a względność (Truth and Relativity)*, Kraków: Aureus.

Grobler, Adam and Wiśniewski, Andrzej, 'Explanation and theory-evaluation' in Roberto Festa *et al.* (eds.) *Volume in Debate with Theo Kuipers*, Kluwer (to appear).

Institute of Philosophy
University of Zielona Góra
Al. Wojska Polskiego 71A
PL-65-762 Zielona Góra
Poland
uzgroble@theta.uj.edu.pl

ILKKA NIINILUOTO

TRUTH-SEEKING BY ABDUCTION

In a seminar with the title "Deduction and Induction in the Sciences", it is intriguing to ask the following questions: (i) Is there a third type of inference besides deduction and induction? (ii) Does this third type of inference play a significant role within scientific inquiry? A positive answer to both of these questions was advocated by Charles S. Peirce throughout his career, even though his opinions changed in important ways during the fifty years between 1865 and 1914. Peirce called the third kind of inference "hypothesis", "abduction", or "retroduction".[1] In this paper, I shall follow Peirce's steps in discussing abduction by analyzing its logical form (Sections 1-3), its role in science (Section 4), and the grounds of its validity (Sections 5-6). We shall see that Peirce's discussion is more insightful than many recent attempts to analyze abductive inference. Still, recently some progress has been made in the treatment of abduction within the Bayesian theory of epistemic probability (Section 6) and truth-approximation (Section 7). The results of this work support the view of scientific realism: abduction or inference to the best explanation (IBE), combined with empirical and experimental testing of scientific theories, is the best method of seeking informative truths in science (Section 8).

1. CLASSIFICATION OF INFERENCES

Peirce introduced the classification of inferences into deduction, induction, and hypothesis in his lectures in 1865-66. While this terminology is taken from Kant, the structure of the argument is derived from Aristotelian syllogistic. These suggestions were later elaborated in the articles 'Deduction, Induction, and Hypothesis' (1878) and 'A Theory of Probable Inference' (1883).

Peirce's starting point is the observation that a paradigm example of *deduction*, the Barbara syllogism of the first figure, can be inverted in two different ways. Using modern logical notation and allowing singular terms, Barbara looks as follows:

(1) $(x)(Fx \rightarrow Gx)$
Fb
$\therefore Gb.$

Induction is the inference of the major premise (rule) from the minor premise (case) and the conclusion (result):

F. Stadler (ed.), Induction and Deduction in the Sciences, 57–82.
© 2004 *Kluwer Academic Publishers.*

(2) Fb
Gb
$\therefore (x)(Fx \rightarrow Gx)$.

Hypothesis is the inference of the minor premise from the major premise and the conclusion:

(3) $(x)(Fx \rightarrow Gx)$
Gb
$\therefore Fb$.

(See *CP* 2.623.) Simple variants of these schemata allow the case to refer to more than one instance b. Then the inversion (2), which Peirce called "crude induction" (*CP* 2.757, 6.473), expresses a typical inductive generalization based on several instances:

(4) $Fb_1 \& \ldots \& Fb_n$
$Gb_1 \& \ldots \& Gb_n$
$\therefore (x)(Fx \rightarrow Gx)$.

The difference between induction and hypothesis was described by Peirce in the following way:

> By induction, we conclude that facts, similar to the observed facts, are true in cases not examined. By hypothesis, we conclude the existence of a fact quite different from anything observed, from which, according to known laws, something observed would necessarily result. The former, is reasoning from particulars to the general law; the latter, from effect to cause. (*CP* 2.536.)

Thus, hypothetical reasoning (3) proceeds from effects to causes, while a typical deduction (1) proceeds from causes to effects. (The latter requirement is Aristotle's well-known condition for scientific syllogisms.) Peirce also stated, already in 1865, that hypothesis is an inference to an *explanation*, and the resulting inference (1) is then an "explanatory syllogism" (*W* 1:267, 428, 440). The same holds of examples where Barbara only involves universal statements:

(5) $(x)(Fx \rightarrow Gx)$
$(x)(Bx \rightarrow Fx)$
$\therefore (x)(Bx \rightarrow Gx)$.

In this case, the hypothetical inversion of (5) takes the form

(6) $(x)(Fx \rightarrow Gx)$
$(x)(Bx \rightarrow Gx)$
$\therefore (x)(Bx \rightarrow Fx)$.

For example, the supposition that light (B) is ether waves (F) explains why light gives certain peculiar fringes (G), given the law that ether waves give these fringes (*W* 1:267).

Hence, Peirce's account of deduction covers what Carl G. Hempel in 1948 called the deductive-nomological (DN) explanation of singular facts (1) and the DN explanation of laws (5). Peirce's account of hypothesis, as an inversion of deduction, covers singular abduction (3) and general abduction (6).[2]

In his 1878 paper, Peirce gave three examples of hypothetical inference (*CP* 2.625). In meeting a Turkish man surrounded by four horsemen holding a canopy over his head, Peirce inferred that this man was the governor of the province. In finding fish-like fossils far in the interior of the country, we suppose that the sea once washed over this land. Countless documents and monuments refer to a conqueror called Napoleon Bonaparte, and we explain them by supposing that Napoleon really existed.

These examples illustrate the fact that inference from effect to cause may involve reasoning backwards in time. This is the motive why in 1898 Peirce called hypothesis *retroduction* (see Peirce, 1992a).[3] As Peirce himself knew, such retroductions are typical in solving crimes both in real life and in detective stories (see Niiniluoto, 1999c).

What Peirce did not observe, however, was that hypothetical inference (proceeding from effects to causes) and explanatory deductions (proceeding from causes to effects) correspond to what medieval and Renaissance Aristotelians called *resolutio* and *compositio*, respectively. These Latin terms were translations of the Greek terms *analysis* and *synthesis*, as used in geometry, so that Peirce's abduction has an interesting historical connection to the geometrical method of analysis (see Niiniluoto, 1999c). This method reasons backwards from theorems to axioms (theoretical analysis) or from solutions to constructions (problematic analysis), while the proof or solution is thereafter given in synthesis.

The example of Napoleon Bonaparte also illustrates the important point that hypothetical reasoning "infers very frequently a fact not capable of direct observation" (*CP* 2.641):

> The great difference between induction and hypothesis is, that the former infers the existence of phenomena such as we have observed in cases which are similar, while hypothesis supposes something of a different kind from what we have directly observed, and frequently something which it would be impossible for us to observe directly. (*CP* 2.640.)

Thus, besides "horizontal" inferences leading to observable facts, hypothetical reasoning is typically "vertical" (cf. Psillos, 1996). In his 1898 Cambridge lectures, Peirce even claimed that inference of a *vera causa*, as a "state of things known to be present", is not really hypothetical but inductive (*EP* 2:54).

To keep the number of different kinds of inferences as three, in his early papers Peirce treated inference by *analogy* as a kind of hypothetical inference: "we find that in certain respects two objects have a strong resemblance, and infer

that they resemble one another strongly in other respects" (*CP* 2.624). In 1883, he formulated a probabilistic argument which clearly involves analogical reasoning (*CP* 2.706). He classified this inference as being a hypothesis, but noted that it has also been treated as "an induction of characters".[4] Peirce illustrated this kind of reasoning with an example where two torn pieces of paper with anonymous writing fit together, and it is inferred that a certain person is the author of these notes (*CP* 2.632). This can be viewed as an inference to a common cause, which explains the coexistence of two apparently separate things (see Salmon, 2000, p. 74).

2. PROBABILISTIC REASONING

Peirce's first formulation of induction and hypothesis as probable arguments in 1867 was inspired by John Venn's frequentist treatment of probability in *The Logic of Chance* (1866). In his 1878 paper, Peirce replaced the universal premise 'All F are G' in (1) by the statistical statement 'Most F are G'. Thus, (1) is replaced by *probable deduction*:

(7) Most of the beans in this bag are white.
These beans are from this bag.
∴ Probably, most of this handful of beans are white.

Its two inversions again lead to induction:

(8) Most of this handful of beans are white.
These beans are from this bag.
∴ Probably, most of the beans in this bag are white

and to hypothesis:

(9) Most of the beans in this bag are white
Most of this handful of beans are white.
∴ Probably, these beans are from this bag.

If induction (8) is rewritten by replacing 'most' with 'few', the inferential pattern can be obtained from (7) by the same transformation (denying the result, accepting the case, and denying the rule) which leads from the Barbara syllogism (1) to the Bocardo syllogism of the third figure:

(10) ~Gb
Fb
∴ (Ex)(Fx & ~Gx).

Induction is thus probabilistic reasoning in the third figure (*CP* 2.627). The transformation (denying the result, accepting the rule, denying the case) leads from Barbara (1) to the Baroco syllogism in the second figure:

(11) ~Gb
(x)(Fx → Gx)
∴~Fb

A similar transformation leads from the probable deduction (7) to

(12) Few beans of this handful are white.
Most beans in this bag are white.
∴ Probably, these beans were taken from another bag.

Thus, hypothetical inference is probabilistic reasoning in the second figure (see also Peirce, 1992a). About the schemata (9) and (12) Peirce made the interesting additional point:

When we adopt a certain hypothesis, it is not alone because it will explain the observed facts, but also because the contrary hypothesis would probably lead to results contrary to the observed. (*CP* 2.628.)

According to another formulation introduced by Gilbert Harman (1968), hypothetical reasoning is not only inference to an explanation but *inference to the best explanation* (IBE).

In his 1883 article, Peirce further formalized probabilistic arguments by replacing the major premise by a precise statistical statement 'The proportion r of Fs are G'. Then we obtain the schema of *simple probable deduction* as a variant of deduction (1):

(13) The proportion r of the Fs are G;
b is an F;
It follows, with probability r, that b is a G.

Here the conclusion can be taken to be 'b is a G', and the probability r indicates "the modality with which this conclusion is drawn and held to be true" (*CP* 2.720). More precisely, the probability of a mode of argument is its *truth-frequency*, i.e., its ability to yield true conclusions from true premises. For example, this applies to the probability r in (13). For ordinary deduction, which is necessarily truth-preserving, the truth-frequency is one.

Peirce generalized (13) to the schema of *Statistical Deduction*:

(14) The proportion r of the Fs are G,
b',b",b"', etc. are a numerous set, taken at random from among the Fs;
Hence, probably and approximately the proportion r of the b's are G.

What Peirce called "quantitative induction" can now be obtained by inverting the argument (14): the premises state the observed proportion of Gs in a random sample of Fs and the conclusion states the proportion of Gs in the whole population of Fs. This inference, later discussed by Hans Reichenbach in his "straight

rule", includes inductive generalization (4) as its special case (r = 1). Another inversion of (13) leads to simple probabilistic abduction:

(15) The proportion r of the Fs are G.
Gb
∴ Fb.

This is a generalization of the singular hypothetical inference (3). However, its truth-frequency q is in general not equal to r, as in (13), since in this inverse inference q depends on the proportion of Fs in the class of Gs, and it is clear that this truth-frequency is independent of r. Indeed, it is also possible that it is close to zero.[5]

Peirce's treatment of probabilistic reasoning anticipated important 20th century frequentist accounts of induction like Reichenbach's probability logic and the Neyman-Pearson theory of statistical tests. If Peirce's truth-frequencies are changed to epistemic probabilities, his model of probable deduction (13) corresponds to what Rudolf Carnap and Isaac Levi have called "direct inference" (see Levi, 1967). This type of inference has often been discussed as a model of prediction. But, in his discussion of probabilistic inference, Peirce repeated the point that hypotheses are adopted "for the sake of the explanation", so that (13) is an explanatory statistical syllogism (*CP* 2.716). Thus, Peirce here anticipated C.G. Hempel's 1962 model of inductive-probabilistic explanation of particular facts. Moreover, his schema for statistical deduction is richer than Hempel's IP model and Salmon's SR model, since it provides an account of probabilistic explanation of statistical facts (see Niiniluoto, 1981). If the major premise of (13) is interpreted as a probabilistic law with a dispositional "would-be" or propensity, as Peirce suggested in 1910 (*CP* 2.666), his schema corresponds to the propensity model of IP explanation (see Fetzer, 1981; Niiniluoto, 1981).

In Niiniluoto (1993), I argued that Wesley Salmon has ignored Peirce's role in the theory of probabilistic explanation. Hempel conceded the point (see Niiniluoto, 2000), but it seems that Salmon remained unconvinced. In his papers (Salmon 2001a,b), which were based on a conference in Haifa in 1998, but appeared after his death on April 2001, Salmon presented a Bayesian critique of inference to the best explanation. Peirce is not mentioned at all. The history of DN explanation is traced to Aristotle, Mill, and Popper, and the history of abduction starts with Harman (1965). When Salmon argues that "we cannot fairly evaluate inference to the best explanation without considering statistical models of explanation" (Salmon, 2001a, p. 72), he is in fact repeating a point made forcefully by Peirce already in 1878.

3. ABDUCTION RECONSIDERED

Peirce's early classification of inferences was based upon Aristotelian syllogisms. However, a complete characterization of logical reasoning cannot be

found in this way, since we are able to give only typical simple examples of different kinds of inferences. Different logical forms of valid deductions cannot be given by any finite list. Peirce realized this quite well when he developed his theory of relatives in the 1870s and his theory of quantification in the 1880s. He did not construct an axiomatic treatment of deduction, as did logicians from Frege and Russell to Hilbert and Gentzen. Instead, he gave a general characterization of validity in semantical terms. In 'On the Algebra of Logic' (1880), Peirce introduced a statement of the form P $\leq$ C to express "the truth of the leading principle" of the inference P $\therefore$ C. This statement signifies that "every state of things in which a proposition of the class P is true is a state of things in which the corresponding propositions of the class C are true" (*EP* 1:203). The characteristic of valid deduction is thus that it is necessarily truth-preserving.

Peirce divided inferences into *analytic* (deductive) and *synthetic* (induction and hypothesis) (*CP* 2.624). Analytic inferences are "explicative", while synthetic inferences are "ampliative". Peirce knew that it is only a "half-truth" to claim with Kant that deduction merely explicates what is implicitly asserted in the premises (*EP* 2:96). Arguing that deduction proceeds by diagrams, Peirce made a distinction between *corollarial* and *theorematic* deduction: the former finds the conclusion by observing the diagram, but the latter also "performs an ingenious experiment upon the diagram" (*EP* 2:298). This distinction corresponds to Jaakko Hintikka's division between trivial and non-trivial deductions (or surface tautologies and depth tautologies), where the latter require the introduction or construction of new individuals in the proof (see Hintikka, 1997).

As an ampliative inference, induction is fallible, but by its self-corrective nature, it is guaranteed to approach the truth in the long run. According to Peirce, induction can be justified by the fact that it "pursues a method which, if duly persisted in, must, in the very nature of things, lead to a result indefinitely approximating to the truth in the long run" (*CP* 2.781).

Peirce's deeper insights about deduction led him to revise his views concerning induction and hypothesis as well. In his 1898 Cambridge lectures, he also called hypothesis *abduction* and *retroduction* (see Peirce, 1992a). Peirce claimed that he had earlier misunderstood the relation between induction and hypothesis (*CP* 8.227). His formulation of the general form of abduction goes as follows:

(16) If μ were true, π, π', π'' would follow as miscellaneous consequences.
But π, π', π'' are in fact true.
$\therefore$ Provisionally, we may suppose that μ is true.

In his 1903 Harvard lectures, Peirce expressed the general form of the "operation of adopting an explanatory hypothesis" as follows (*CP* 5.189, *EP* 2:231):

(17) The surprising fact C is observed;
But if A were true, C would be a matter of course.
Hence, there is reason to suspect that A is true.

This schema indicates how a hypothesis can be "abductively conjectured" if it accounts "for the facts or some of them". As in this schema (17) A may be a general theory (in contrast to singular abduction (3)), it might be said to express *theoretical abduction*. Moreover, Peirce added, the conclusion is not A itself, but the assertion that "there is reason to suspect that A is true". In this sense, retroduction "only infers a *may-be*" from an actual fact (*CP* 8.238).

Already in 1878 Peirce argued that hypotheses have to be put into "fair and unbiased" tests by comparing their predictions with observations (*CP* 2.634). In his papers and lectures in 1901-03, Peirce defined induction in a new way as "the operation of testing a hypothesis by experiment" (*CP* 6.526), whereas abduction is an "inferential step" which is "the first starting of a hypothesis and the entertaining of it, whether as a simple interrogation or with any degree of confidence" (*CP* 6.525). Here abduction and induction are successive steps in scientific inquiry:

> Abduction, on the other hand, is merely preparatory. It is the first step of scientific reasoning, as induction is the concluding step... The method of either is the very reverse of the other's. Abduction makes its start from the facts, without, at the outset, having any particular theory in view, though it is motivated by the feeling that a theory is needed to explain the surprising facts. Induction makes its start from a hypothesis which seems to recommend itself, without at the outset having any particular facts in view, though it feels the need of facts to support the theory. Abduction seeks a theory. Induction seeks for facts. (*EP* 2:106.)

> Abduction is the process of forming an explanatory hypothesis. It is the only logical operation which introduces any new idea; for induction does nothing but determine a value and deduction merely evolves the necessary consequences of a pure hypothesis.
>
> Deduction proves that something *must* be, Induction shows that something *actually is* operative, Abduction merely suggests that something *may be*.
>
> Its only justification is that from its suggestion deduction can draw a prediction which can be tested by induction and that, if we are ever to learn anything or to understand phenomena at all, it must be by abduction that this is to be brought about. (*EP* 2:216.)

Peirce's account is in harmony with the method of hypothesis, as advocated earlier by William Whewell and later by Karl Popper (see Niiniluoto, 1984), but Peirce, like Whewell, admitted that there is an inference from our original problem situation to a hypothesis.[6] After the abductive step in proposing the hypothesis A, the next step is to derive consequences from A by deduction, and then to put theory A into a severe observational or experimental test (*EP* 2:114). This follows the hypothetico-deductive (HD) model of science, but Peirce also allowed for cases where the test evidence is only a probabilistic consequence of the hypothesis.[7] A limiting case of abduction is perception (*CP* 5.181, *EP* 2:227), but abduction may lead backwards to historical causes and to general scientific theories as well. Peirce argued sharply against Auguste Comte that science should not be restricted to hypotheses "verifiable by direct observation", since such a restriction would "cut off all history as an admissible hypothesis" (*EP* 2:225).

4. TRUTH-SEEKING: DISCOVERY AND PURSUIT

What role could an "inference to an explanation" have in science? One answer follows the old Aristotelian "regressive" idea that science wishes to proceed from *knowing that* to *knowing why*: once we know that grass is green, we would like to have a deeper insight into the causes of this fact, i.e., we wish to know why grass is green. This desire is fulfilled by an explanation of the fact that grass is green. A similar distinction was made by Hempel (1965) between *reason-seeking* why-questions (what reasons do we have for thinking or claiming that p?) and *explanation-seeking* why-questions (why is it the case that p?). In the logic of questions, the statement that p is taken to be a presupposition for the question 'why p?'.

Another motive in proposing explanations is to advance our knowledge from its present state to a higher stage: if we already know that p, and a hypothesis q explains why p, then q is a good candidate for being true as well. In this view, finding the best explanation of known facts and regularities is a fallible or non-demonstrative method of *truth-seeking*. This attempt to find new truths is clearly the main function of Peirce's abduction. In other words, abduction has a crucial role in learning (Nesher, 2001).

Salmon (2001a, p. 79) argued that the characterization of non-demonstrative inference as inference to the best explanation serves "to muddy the waters" by fostering confusion between Hempel's two types of why-questions. However, if we infer q on the basis of its ability to explain a known fact p, we are indeed explaining why p in Hempel's sense, but we are not thereby proposing new reasons for believing that p, but rather seeking abductive reasons for believing that q. Similarly, reasons for believing that q should be distinguished from the further question 'Why q?'. Hence, no confusion between Hempel's two questions is involved here.

The role of abduction has often been interpreted as the heuristic function of the *discovery* of new theories, or alternatively as the motive for suggesting or *pursuing* testworthy hypotheses (see Hanson, 1961).[8] This is in harmony with Peirce's later view of abduction, deduction, and induction as successive steps in inquiry. In this sense, abduction plays an important *strategic* role in truth-seeking (cf. Hintikka, 1998; Schurz, 2003). Proposing theoretical hypotheses at random is completely ineffective as a means to find new truths. Inquiry must start from "surprise", i.e., from a problem to be solved, and our first step is to restrict our attention to those hypotheses which are capable of solving our initial problem. If the problem is expressible as a why- or how-question, we should strategically limit our attention to "a problem space" consisting of hypotheses that have explanatory power with respect to the initial question. This is the basic rule of a "data-driven" logic of discovery (cf. Langley et al., 1987). In the next steps, we may look for further reasons or constraints to narrow down the space of search.

In retroduction, the search process is similar to geometrical analysis: a causal chain is traced by starting from its end and by reasoning backwards to earlier causes with the help of additional questions (cf. Niiniluoto, 1999c; Schurz, 2003). Theoretical abduction can be guided by background knowledge (e.g., the mathematical form of the desired law may be inferred from earlier theories) or by simplicity considerations (e.g., the relevant laws are classified and studied by following a simplicity ordering) (cf. Langley et al., 1987; Niiniluoto, 1994). It is also possible that the surprising observations are anomalous with the best theory so far (see Aliseda-Llera, 1997; Kuipers, 1999), in which case what we are searching for is an explanation of the new data, which at the same time is an improvement over, or a revision of, the earlier theory.

In the same way, making experiments at random is ineffective as an inductive method, at least if we do not know what we are seeking. Peirce was familiar with the idea that good experiments are "questions put to nature" (see *EP* 2:215), i.e., they have to be designed so that they give answers to our requests of important information. In his interrogative model of inquiry, Hintikka (1998) has suggested that abduction and induction could be understood as "interrogative moves" where new information is obtained by presenting questions to an oracle (such as nature or some other source), so that in a sense the only type of inference is logical deduction. This framework of questioning is powerful for studying the strategic function ("force" in Peirce's terms) of inference (see also *EP* 2:109). However, in genuine discoveries the required information is not yet available anywhere, but has to be created by the researcher herself. Furthermore, it is also possible to extend Hintikka's interrogative model to cases where the reliability of the oracles vary (cf. Niiniluoto, 1997), so that accepting their answers involves non-demonstrative steps of reasoning.

Let us illustrate these ideas with current research in high energy physics. After the success of atomic theory in the first half of the 20th century and its experimental implementation in nuclear weapons in the 1940s and nuclear power plants in the 1950s, an attempt at a deeper level of understanding was made by the quark theory of the 1960s. According to what is now known as the Standard Theory, all matter in the universe is made out of four kinds of building blocks: electrons, neutrinos, up-quarks, and down-quarks. The atomic nucleus is formed by protons and neutrons, which are constituted by up-quarks and down-quarks, glued together by the strong force carried by gluons. The electromagnetic force is carried by photons, gravity by gravitons, and the weak force by W and Z particles. When the universe was born in the Big Bang, there still existed antimatter besides matter. Gravitons have not yet been discovered, but W and Z particles were found in experiments in 1982 at the European Laboratory for Particle Physics CERN.

An important addition to the Standard Theory was introduced by Higgs around 1970. In order to explain why particles have mass, Higgs postulated the existence of a field and argued that mass is formed through the interaction of a particle with this field. From quantum theory and Einstein's equation $E = mc^2$ it

follows that the Higgs field should be associated with (at least one) Higgs particle.

The physicists know that the Standard Theory is still incomplete, but they are vigorously pursuing its completion. Theoretical derivations are made in order to guarantee that the theory is able to account for the accepted picture of existing particles and forces. An important argument in favor of the Standard Theory is that the Higgs particle explains all other particles, while other theories fail for some known particles. In other words, the Standard Theory is worthy of further investigation, since it has more explanatory power than its rivals. To this abductive argument it is often added that this theory is also fertile in the sense that it serves as a starting point for unified field theories (such as string theory).

The pursuit of the Standard Theory also involves gigantic efforts in testing its correctness. Theoretical calculations are made in order to find testable consequences of this theory. Such tests can be made by producing beams of particles, letting them collide, detecting the results, and analyzing the data by computers. As the production of heavy particles requires high energy levels, the design of relevant experiments and the constructions of needed equipment require advanced and expensive technology. The most important open problem is the existence of the Higgs particle. The CERN Laboratory is now replacing its Large Electron Positron collider (LEP), which resides in a circular tunnel of the length of 27 kilometers, with a new accelerator, the Large Hadron Collider (LHC). When its activity starts in 2007, more than 4000 scientists will be collaborating in the hunt for the Higgs particle. The experiment is expected to run for a decade. If it turns out to be successful, the Standard Theory will be confirmed and accepted by the physics community; otherwise, alternative theories about the constitution of matter have to be investigated.

5. ABDUCTION AND CONFIRMATION

We have seen in Section 4 that abductive arguments are powerful enough to convince governments to spend billions of dollars and euros to pursue programs of experimental high energy physics. But no agreement has been reached by the philosophers about the epistemological significance of inferences appealing to the explanatory virtues of theories.

One way of understanding the role of abduction is to claim that it belongs entirely in the context of discovery, while the context of *justification* includes the process of testing a hypothesis by empirical evidence. According to the standard view about the HD method, negative test results disconfirm or refute a general theory, positive results *confirm* the hypothetical theory or give *inductive support* to it. A stronger notion of justification involves the notion of *acceptance*: a hypothesis is so strongly supported by evidence that it is rational to tentatively accept it as true, i.e., to include it in the body of what is currently taken to be scientific knowledge.

Another way of claiming that abduction has no role in the context of justification is to emphasize the distinction between explanation and prediction and to include only predictive success in the context of confirmation (see Salmon, 2000a). In the same way, it is sometimes claimed that *accommodation* (i.e., the adoption of an explanatory hypothesis to account for already known data) has to be sharply distinguished from *prediction* in the sense that accommodation does not have any power of confirmation.[9]

From the strategic viewpoint one might also suggest that the separate evaluation of specific abductive inferences is unnecessary. Instead, the relevant question is whether such steps contribute to the overall or ultimate goal of inquiry. In this sense, Peirce himself defended abduction as the only form of inference that introduces new ideas. Abduction as such is "nothing but guessing" (*EP* 2: 107). The justification of the abductively conjectured hypothesis, which is a "may-be" or a possible truth, is then left to the process of testing its empirical predictions.

However, Peirce himself pointed out that in some cases abductive inference is "compelling" (e.g., perception in ordinary circumstances) or "perfectly certain" (e.g., the existence of Napoleon Bonaparte) (*EP* 2:54). Therefore, it seems appropriate to study specific abductive arguments such as (3), (6), (9), or (15). As we have already observed, the Peircean truth-frequencies of such inferences may depend on the context, but that does not exclude the possibility of an interesting theory about such inferences. By comparison, here we may refer to the enormous literature studying the validity of ampliative arguments of the form (4). The attempt to associate such inductive generalizations with epistemic probabilities within inductive logic has led to the general conclusion that such inductive probabilities depend not only on the premises and the conclusion, but also on some contextual factors or presuppositions (cf. Niiniluoto and Tuomela, 1973). Even though Peirce as a frequentist was sharply critical of the classical Bayesian theory of probabilistic inference (see e.g. *EP* 2:215), we shall see that Bayesianism provides a useful framework for studying abduction and induction as forms of ampliative reasoning.

Let us start with the qualitative concept of confirmation. Howard Smokler (1968) proposed that "abductive inference" satisfies the following principles of Converse Entailment and Converse Consequence:

(CE) If hypothesis H logically entails evidence E, then E confirms H.

(CC) If K logically entails H and E confirms H, then E confirms K.

A variant of CE is obtained by replacing deduction with explanation (see Niiniluoto and Tuomela, 1973):

(CE*) If hypothesis H deductively explains evidence E, then E confirms H.

Smokler contrasted these principles with conditions of Entailment and Special Consequence:

(E) If evidence E logically entails hypothesis H, then E confirms H.

(SC) If E confirms H and H logically entails K, then E confirms K.

These principles are satisfied by "enumerative inference". For example, if all ravens observed so far support the generalization that all ravens are black, they also support the consequence that the next raven is black. Hempel's (1965) proof that these abductive and enumerative principles of confirmation cannot reasonably be satisfied together gives further motivation, in a sense, to Peirce's original thesis that abduction and induction are two different kinds of ampliative inference.

Let us move on to comparative and quantitative notions of confirmation. Assume that a probability measure P is available for the relevant language. In the classical theory of probability (Laplace, Jevons), Bayes's Theorem was used for the study of "inverse probabilities", i.e., the probability of alternative causes H_i given an effect E. For example, E may be (a statement describing) the event of a person's death, and H_i (i =1, ..., n) describe the possible causes of her death. Given statistical mortality tables, let the prior probability $p_i = P(H_i)$ be the relative frequency of the occurrence of event H_i, and let the likelihood $q_i = P(E/H_i)$ be the probability that H_i produces the effect E. Then Bayes's Theorem gives the posterior probability of the cause H_i given E:

(18) $P(H_i/E) = p_i\, q_i / P(E)$,

where

(19) $$P(E) = \sum_{j=1}^{n} p_j\, q_j\,.$$

Such posterior probabilities (18) are "abductive probabilities" in a Peircean sense since they proceed from effects to causes, as it were.

The same formula (18) can be used when H_i is a hypothetical theory. In this case, the probabilities of statements H_i and E are usually interpreted as *epistemic probabilities*: $P(H_i)$ is our initial degree of belief in the truth of H_i, and $P(H_i/E)$ is this degree of belief given evidence E. Note that the application of Bayes's Theorem presupposes that the hypotheses H_i constitute an "ultimate partition" (Levi, 1969) of mutually exclusive and jointly exhaustive statements.

The principles (CE) and (CE*) are satisfied if confirmation is defined by the *Positive Relevance* criterion: E confirms H if and only if P(H/E) > P(H). According to Bayes's Theorem (18),

(20) $P(H/E) = P(H)P(E/H)/P(E)$.

Now if H entails E, we have P(E/H) = 1. Hence,

(21) If H logically entails E, and if P(H) > 0 and P(E) < 1, then P(H/E) > P(H).

More generally, as positive relevance is a symmetrical relation, it is sufficient for the confirmation of H by E that H is positively relevant to E. If inductive explanation is defined by the positive relevance condition, i.e., H inductively explains E if and only if P(E/H) > P(E) (see Niiniluoto and Tuomela, 1973; Festa, 1999), then we have the general result:

(22) If H deductively or inductively explains E, then E confirms H.

The "enumerative" principles E and SC are instead satisfied by the *High Probability* criterion: E HP-confirms H if and only if P(H/E) is sufficiently large (i.e., at least as large as a fixed threshold value). If both explanation and confirmation are defined by the high probability criterion, which is not symmetrical, then the result (22) does not hold any more.

Salmon's recent papers challenge the "explanationist" view that explanatory virtues are relevant to confirmation: "confirmation is logically independent of explanation" (Salmon, 2001a, p. 88). Salmon's claim is that we should first find the most probable of the rival explanations and only then use it as an explanation of known facts (*ibid.*, p. 83). Even after his discussion with Peter Lipton, Salmon remained a "bashful Bayesian" who is "reluctant to become too friendly with Inference to the Best Explanation" (Salmon, 2001b, p. 126).

Lipton (1991) distinguishes "likely" explanations, which take into account the overall credibility or warrant of the explanation, and "lovely" explanations, which provide potential understanding about the explanandum. He argues further that loveliness is at least a guide to likeliness. He concludes (Lipton, 2001b) that likeliness can be defined by posterior probability P(H/E), whereas his notion of "loveliness" might correspond to explanations with a high value of P(E/H) (cf. Niiniluoto, 1999b). Okasha (2000) interprets Lipton as saying that P(H) indicates likeliness and P(E/H) loveliness. Salmon (2001b) also considers the role of prior probabilities and likelihoods as possible explanatory virtues. However, the discussion remains largely inconclusive, since Salmon and Lipton do not appeal to any formal measures of explanatory power. They also seem to take it for granted that confirmation has to be explicated by high posterior probability – even though Salmon mentions the Popperian idea that scientists often prefer bolder hypotheses because of their informational value at the expense of their probability (*ibid.*, p. 121).

A comparative version of (22) would claim that better explanations receive better confirmation:

(23) If K is a better explanation of E than H, then E confirms K more than H.

To prove this result, we need a quantitative definition of the *explanatory power* expl(H,E) of H with respect to E, and of the *degree of confirmation* conf(H,E) of H by E. Popper (1959) attempted to achieve this directly by defining his measure of corroboration as a function of explanatory power. If we define

$$expl_1(H,E) = [P(E/H) - P(E)]/[1 - P(E)]$$

(see Hintikka, 1968; Niiniluoto and Tuomela, 1973), then K is a better explanation of E than H if and only if $P(E/K) > P(E/H)$. If now degree of confirmation is defined by the ratio measure:

$$conf_1(H/E) = P(E/H)/P(E)$$

(see Festa, 1999; Kuipers, 2000), result (23) holds.

If explanatory power is defined by Hempel's 1948 measure

$$expl_2(H,E) = P(\sim H/\sim E) = [1 - P(H \vee E]/[1 - P(E)]$$

(see Hempel, 1965), then K is a better explanation of E than H if and only if $P(K)(1 - P(E/K)) < P(H)(1 - P(E/H))$. This measure thereby favors theories with high information content $cont(K) = 1 - P(K)$ and high likelihood $P(E/K)$. But if better explanation is defined by the expected value of $expl_2(H,E)$ (see Niiniluoto, 1999a, p. 187), then this comparative notion is correlated with the difference measure of confirmation:

$$conf_2(H/E) = P(H/E) - P(H) = P(H)[P(E/H) - P(E)]/P(E).$$

For many purposes, $conf_2(H/E)$ gives similar results as the posterior probability measure

$$conf_3(H/E) = P(H/E).$$

For example, we see that higher initial plausibility and higher explanatory likelihood together are sufficient for higher confirmation:

(24) If $P(K) \geq P(H)$ and $P(E/K) \geq P(E/H) > P(E)$,
then $conf_2(K/E) \geq conf_2(H/E)$ and $conf_3(K/E) \geq conf_3(K/E)$.

(Cf. Okasha, 2000.) On the other hand, there is an important difference between $conf_2$ and $conf_3$. Suppose that K is logically stronger than H, i.e., K entails H. Suppose also that K explains something (a part of the evidence E) that is not explained by H. Then K has greater explanatory power than H, but still K is less probable than H, i.e., $P(K) < P(H)$, and $P(K/E) < P(H/E)$ (see Monton, 1998). (This shows that Lipton's loveliness is not always a guide to his likeliness.) However, in this situation it must be the case that $P(E/K) > P(E/H)$, so that we may have $conf_2(K/E) > conf_2(H/E)$.

So far we have only discussed the relation of explanation and confirmation. We have argued that the Bayesian model of inference helps to show how evidence may confirm hypotheses that are abductively introduced to explain them. However, the formal results about confirmation can be immediately generalized to more complex situations. First, all the probabilities can be relativized to background knowledge B. Secondly, besides the initial evidence E defining our original problem situation, there may additional "old evidence" E_0, and it may happen that the hypothesis H proposed on the basis of E turns out to explain E_0 as well.[10]

Then, using the measure $conf_2$, the confirmation of H by E and E_0 together is defined by $conf_2(H,E\&E_0)$, and the confirmation of H by E_0 relative to E by

$$conf_2(H,E_0/E) = P(H/E\&E_0) - P(H/E).^{11}$$

Thirdly, degrees of confirmation can be defined by similar formulas when new evidence E' is obtained, i.e., the evidence E' is the result of a test, performed after the adoption of H as a hypothesis from which E' can be logically or probabilistically derived. In this way we obtain a probabilistic model of the method of hypothesis, eloquently discussed by Peirce in his later work. The prior plausibility of a hypothesis is now its probability P(H/E) relative to the initial evidence E, and depends on the ability of H to explain E. The confirmation of H due to the new evidence E' is then definable by $conf_2(H,E'/E)$. Again, the joint confirmation of H by E and E' can be defined by $conf_2(H/E\&E')$. In typical cases, the novel facts E' that H has predicted become, after the verification of E', parts of the total evidence that H explains. Therefore, it seems that we need not assume an essential difference between accommodation and prediction as methods of confirmation (cf. Howson and Urbach, 1989). Similarly, the measures of explanatory power can be reinterpreted so that they cover both explanatory and predictive relations between hypothesis and evidence. Hempel (1965) used the notion of *systematic power* as the combination of the explanatory and predictive power of a hypothesis relative to the total evidence E. Then the degree of confirmation of a hypothesis H on the total evidence E depends on the explanatory and predictive power of H with respect to E. For this reason, most discussions about abduction within the Bayesian framework do not distinguish between explanatory and predictive success: what they are really concerned with are thus the relations between systematic power and confirmation.

6. Abduction and Acceptance

The relation of better explanation also defines the notion of the *best* explanation within a set of rival hypotheses. Harman (1965) formulates *inference to the best explanation* by the following rule:

(IBE) A hypothesis H may be inferred from evidence E when H is a better explanation of E than any other rival hypothesis.

Here, inference is understood in the strong sense of acceptance. IBE thus claims that the best explanation H of evidence E is inductively acceptable on E. Harman also tries to reduce all inductive inference to IBE, but this is problematic already for inductive generalization, since in schema (4) the conclusion alone does not explain the premises.

Different explications of explanatory power now lead to alternative versions of IBE. Measure $expl_1(H,E)$ implies the maximum likelihood principle, familiar from statistical inference:

(25) Given evidence E, accept the theory H that maximizes the likelihood P(E/H).

The expected value of $expl_2(H,E)$ leads to

(26) Given evidence E, accept the explanation H of E such that H maximizes the difference P(H/E) - P(H).

The high probability rule recommends the acceptance of the most probable explanation:

(27) Given evidence E, accept the explanation H of E such that H has the maximal posterior probability P(H/E) on E.

As self-explanations are excluded, in (27) H cannot be E (or contain E as its conjunct), even though P(E/E) has the maximal value 1. For theories which deductively explain E, it follows from Bayes's Theorem (20) that rule (27) recommends the choice of the theory H that has the highest prior probability P(H), i.e., H is initially the most plausible of the rival explanations. With the same assumptions, this H also maximizes the difference P(H/E) - P(H), so that (26) and (27) are equivalent in this case. However, if rule (27) is applied to inductive explanation as well, (25) and (27) may lead to different results: the theory with maximal likelihood need not be the one with the highest posterior probability, if it has a very low prior probability.

We argued above that explanation and prediction behave in a symmetrical fashion with respect to the probabilistic confirmation of rival hypotheses. However, the case of acceptance can be argued to be different. A hypothesis that explains our initial data, and is thereby confirmed by it to some extent, may still be ad hoc. To remove this doubt, the hypothesis should either explain some old evidence or it should be successful in serious new tests. As we have seen, this is what Peirce (following Whewell) demanded. Therefore, one may argue that IBE as an acceptance rule should contain an additional condition stating that the "best" hypothesis is one with both explanatory and predictive power. For example, the Standard Theory will not be accepted by its great explanatory power alone, but it has to turn out to be successful in experimental tests.

As a rule of acceptance, IBE also faces the problem that sometimes the best explanation is not good enough for being a serious candidate for truth (see Salmon, 2001a, pp. 85-86), even though (as we have seen) the evidence gives some confirmation of it. Lipton (2001b, p. 104) suggests that IBE should be formulated as "inference to the best of the available competing explanations, when the best one is sufficiently good". The same situation holds even in cases when new test evidence is obtained. If a Bayesian account of acceptance allows for the strongest alternative to be a disjunction of the members of an ultimate partition (cf. Levi, 1967; Niiniluoto, 1987), then in some cases *suspension of judgment* is the most rational attitude towards rival hypotheses. In practice, such suspension of judgment can be interpreted as the recommendation that the rival

explanations are still pursued – developed theoretically and tested experimentally – until one of them is deemed good enough for tentative acceptance.

Bas van Fraassen's (1980, 1989) much discussed criticism of abduction concerns IBE as a rule of acceptance, i.e., as a rule that "selects" the best among a set of proposed explanations.[12] His discussion of the possibility of a Bayesian analysis of IBE operates with the notion of belief related to high posterior probability, so that his target can be taken to be (27). However, his three arguments, if they were valid, would also serve to refute the connection between abduction and confirmation.

Peirce repeatedly appealed to Galileo's notion of *il lumen naturale*, our natural disposition to light upon the correct hypothesis in a finite number of guesses (*CP* 7.223). According to Peirce, "man has a certain Insight, not strong enough to be oftener right than wrong, but strong enough not to be overwhelmingly more often wrong than right" (*EP* 2:217). But van Fraassen's "bad lot" argument claims that IBE does not satisfy even the minimum condition that an acceptable hypothesis should have at least the probability 1/2:

> For me to take it that the best of set X will be more likely to be true than not, requires a prior belief that the truth is already more likely to be found in X, than not. (Van Fraassen, 1989, p. 143.)

A simple Bayesian reply to this (see Niiniluoto, 1999b) is to point out that Bayes's Theorem (18) has to be applied to an ultimate partition, so that the disjunction of hypotheses $H_1, \ldots, H_n$ is a logical truth and, hence, precisely one of them is true. Another reply is that van Fraassen's quoted claim is in fact incorrect. To see this, assume that $n = 3$, H_1 deductively explains E, and H_2 inductively explains E, but $p_1 + p_2 < .5$, so that our prior belief is that the truth is not to be found in the set $\{H_1, H_2\}$. For example, let $p_1 = p_2 = .2$, $p_3 = .6$, $q_1 = 1$, $q_2 = .6$, and $q_3 = .1$. Then the best explanation of E is H_1. By (20), $P(E) = .38$, and the posterior probability of H_1 on E is .53.

The application of (19) presupposes that the probability of evidence E on all alternative hypotheses H_i is known. If H_1 is the explanation of E, then we should know $P(E/\sim H_1)$, which is the weighted average of the likelihoods $P(E/H_i)$ for $i \neq 1$. It was argued already by George Boole in 1854 that this assumption is unrealistic (see Boole, 1958, p. 367). However, if $P(E/H) = 1$ and if we allow $P(E/\sim H)$ to take any values in the open interval (0,1), then by (20) P(H/E) takes values in the interval (P(H),1). This is another way stating the result (21): successful deductive explanation of E by H increases the probability of H on E, even when the likelihood $P(E/\sim H)$ is unknown. If this likelihood is sufficiently small, so that H is the only explanation of E, then the increase in the probability P(H/E) will be significant.

Van Fraassen's second criticism claims that the Bayesian account of abduction leads to incoherence: if explanatory success is credited with adding an extra bonus to the posterior probability, this procedure allows the construction of a Dutch Book against the scientist (*ibid.*, p. 169). However, the point of the

Bayesian approach is not to add separate bonus points, but to show that explanatory virtues enter into the assessment of posterior probability via the normal elements of Bayes's Theorem, viz. prior probability and likelihood. No such special assumption was needed to prove the basic result (21) about the connection between explanation and confirmation.

Van Fraassen's third criticism notes that the result (21) about confirmation is blocked if the prior probability P(H) is zero. In the next Section, we shall discuss two cases in which such a zero probability is natural: (a) H is a *sharp* hypothesis, i.e., H belongs to an infinite set of rival hypotheses H, e.g., the set of real numbers; (b) H is an *idealized* hypothesis, i.e., the class of rival hypotheses (including H) is defined relative to some counterfactual assumption (cf. Niiniluoto, 1999a). We shall then return to a criticism of van Fraassen's constructive empiricism in Section 8.

7. IBE AND TRUTHLIKENESS

In this section, some generalizations of abductive inference will be mentioned.[13] Besides deductive and inductive explanation, the analysis of the notion of "the best explanation" should also cover *approximate explanations*: H approximately explains E if and only if E' which is close to E is deducible from H. Alternatively, H' which is close to H entails E. For example, in the problem of curve-fitting the original observational data E is usually incompatible with the considered hypotheses H, so that P(E/H&B) = 0. Similarly, the consequences of highly abstract theories are usually only approximately verified in experiments. Such a discrepancy between H and E may be due to the fact that observational errors make evidence statement E incorrect, so that theory H offers a *corrective explanation* of E. Another possibility is that hypothesis H is an *idealization* involving some counterfactual assumption. For such cases, the probability P(E/H&B) has to be replaced by a measure of *similarity* or *fit* between E and H (see Niiniluoto, 1994). It is suggested in Niiniluoto (1999b) that here the evidence may still indicate that the best hypothesis is *truthlike*. This principle might be called *inference to the best approximate explanation*:

(IBAE) If the best available explanation H of evidence E is approximate, conclude for the time being that H is truthlike.

A variant of IBAE replaces truthlikeness by the weaker notion of approximate truth:

(IBAE') If the best available explanation H of evidence E is approximate, conclude for the time being that H is approximately true.

If degrees of truthlikeness are introduced (see Niiniluoto, 1987, 1998), then the greater the fit between H and E, the larger the degree of truthlikeness of H in the conclusion of IBAE. Already in 1903, Peirce proposed that the observed agree-

ment between predictions and perceptions is to be taken as "the provisional and approximate, or probametric, measure of the general agreement of the theory with facts" (*EP* 2:225).

It is important to note here that the probabilistic account of IBE given by the results (22) and (23) is not directly applicable to rules like IBAE. These results establish *a probabilistic link between explanatory power and truth*: posterior probability P(H/E) is the rational degree of belief in the truth of H on the basis of E, and thereby confirmation (as increase of probability by new evidence) means that we rationally become more certain of the truth of H than before. But a rule of the form IBAE needs *a link between approximate explanation and truthlikeness*. New tools are needed to make sense of this link, since the approximate explanation of E by H allows for H to be inconsistent with E, so that P(E/H) and P(H/E) are zero.

One important tool is to define the notion of *probable approximate truth* (see Niiniluoto, 1987, p. 280). A theory is approximately true if it allows states of affairs that are close to the truth. Hence, the probability PA(H) that a theory H is approximately true (within a specified degree δ) may be greater than zero even though P(H) = 0. This helps us to give a reply to van Fraassen (1989), who argues that the result (21) does not apply to hypotheses H with a zero prior probability (see Niiniluoto, 1999a, p. 188). Note first that Bayesian statistics treats hypotheses with zero probability by introducing a prior probability distribution on the set of hypotheses H and applying Bayes's Theorem for distributions. To accept a sharp null hypothesis H then is often taken to mean that we behave (make decisions) *as if* H were true. However, the notion of probable approximate truth allows us to assess the cognitive virtue of H in spite of its zero probability. The ordinary notion of confirmation as positive probabilistic relevance can be modified by replacing P with PA: let us say that E *ap-confirms* H if PA(H/E) > PA(H) (see Festa, 1999). Let H^δ be the disjunction of H with states of affairs that are at most within the distance δ from some state of affairs allowed by H. Hence, H is approximately true if and only if H^δ is true. Further, PA(H) > 0 if and only if $P(H^\delta) > 0$. Then (21) can be generalized to the following result:

(28) If PA(H) > 0, and H^δ logically entails E, then E ap-confirms H.

(See Niiniluoto, 1999a, p. 188.) Note, however, that this kind of result does not yet justify IBAE', since in this case H is compatible with E.

Another tool in connecting objective degrees of truthlikeness and epistemic matters is based on the idea of estimating verisimilitude by the expected degree of truthlikeness (see Niiniluoto, 1987, p. 269). If $C_1, \ldots, C_m$ are the alternative complete states of affairs expressible in some language, and the degree of truthlikeness of theory H would be $Tr(H,C_i)$ if C_i were the true state, then the *expected verisimilitude* ver(H/E) of H given evidence E is defined by

(29) $$ver(H/E) = \sum P(C_i/E)\, Tr(H,C_i),$$

where the sum goes over all i = 1, ..., m. Function ver estimates truthlikeness in the following sense: if evidence E is true and in the limit gives exhaustive information about the alternatives C_1, ..., C_m, then ver(H/E) approaches the degree of truthlikeness Tr(H,C*) of H relative to the true state C* (see *ibid.*, p. 276). The value of ver(H/E) generally differs from the probability P(H/E). Again we may generalize the notion of probabilistic confirmation: let us say that E *ver-confirms* H (relative to background assumption B) if and only if ver(H/E&B) > ver(H/B). Then, for example, E ver-confirms H if H is the only deductive explanation of E (see Niiniluoto, 1999a, p. 186).

The measure of expected truthlikeness has the important feature that ver(H/E) may be non-zero, and even high, even though E refutes H and P(H/E) = 0. Now we may study the conditions under which the following kinds of principles are valid:

(30) If H' is a better approximate explanation of E than H,
then ver(H'/E) > ver(H/E).

(31) If H approximately explains E,
then the expected verisimilitude of H given E is high.

What the estimated truthlikeness function ver hopefully does in such principles is establish a sort of fallible but rational bridge between the approximate explanatory success of H and the truthlikeness of H. This is a generalization of the probabilistic bridge established between explanation and truth in (21).

If all the relevant hypotheses are idealizations (i.e., defined relative to some counterfactual assumption B), then ver (H/E) has to be calculated indirectly. One way of doing so is by replacing E with E(B), which is what E would have been if B were true. With this modification, principles (30) and (31) can be applied to situations involving idealized hypotheses (see Niiniluoto, 1987, pp. 285-288; 2002).

8. CONCLUSION: ABDUCTION AND SCIENTIFIC REALISM

Scientific realism claims that theories in natural science are attempts to describe a mind-independent reality. Such theories explain and predict observable phenomena, but their content refers to entities and processes beyond the observable domain. While naive realists assert that successful theories are literally true, *critical scientific realists* hold that even the best theories are truthlike or approximately true (see Niiniluoto, 1999a).

In his well-known "confutation of scientific realism", Larry Laudan (1984) demanded that the realists show there to be an "upward path" from the empirical success of science to the approximate truth of theories – and then a "downward path" from approximate truth to empirical success. In my own work, I have tried

to reply to Laudan's challenge by using the concept of truthlikeness and its estimation by function ver (see Niiniluoto, 1984, Ch. 7; 1987).[14]

Many attempts to defend scientific realism by the general "no miracle argument" appeal to forms of abduction which conclude that successful scientific theories are approximately true (see Psillos, 1999, p. 78). In other words, they involve something like the principle IBAE'. Similar arguments can be applied to specific scientific theories, like Newton's mechanics which is able to give approximately correct explanations and predictions of the behavior of ordinary macroscopic physical objects. To infer that the empirical success of Newton's theory can be explained by the hypothesis that this theory is in fact truthlike or approximately true, one needs to defend abduction in the form of IBAE or IBAE' (see Section 7). Our discussion was in principle completely general in the sense that no assumption about the nature of the explanatory hypothesis H was made: the arguments establishing fallible links between empirical success and truthlikeness were entirely independent of the question of whether H involves theoretical terms or not.

Bas van Fraassen's alternative to scientific realism is his "constructive empiricism": instead of accepting theories as *true*, the proper aim of science is *empirical adequacy*, i.e., in accepting a theory H we are only committed to believing that all of its observable consequences are true. Theory H may be treated *as if* it were true (see Ladyman *et al.*, 1997), but in fact one is only interested in what H says about the observable. Therefore, van Fraassen has to deny the validity of abductive arguments from empirical success to the truth of a theory.

It has been pointed out against van Fraassen that his argument against abduction is incoherent, since he has to accept abductive inferences in so far as they are confined to the observable realm: "if one has no reason to doubt commonsense abductive reasoning, then one should have no reason to doubt abduction in science" (Psillos, 1999, p. 211). This is denied by Ladyman *et al.* (1997), who urge that van Fraassen rejects abduction in all cases. But recall that the notion of empirical adequacy employs the notion of truth for observational statements. The rejection of abduction with respect to the observational language seems to commit van Fraassen to a phenomenalist interpretation of that language. A strong argument in this direction can be based upon Peirce's thesis that perception is already a limiting case of abduction: if we assume something like a causal theory of perception, then perceptual impressions in us are caused by external objects, and perceptual judgments thus arise by abductive inferences from effects to causes. In other words, statements about observable properties of observable objects (in what was called the "physicalist" language by logical empiricists in the 1930s), and thereby the foundation of empirical science, rest on abduction. To avoid this conclusion, van Fraassen's constructive empiricist should adopt some sort of phenomenalism.

Stathis Psillos (1996) has attacked van Fraassen by claiming that inference from empirical success to empirical adequacy involves abduction. I think this is correct in the sense that such an inference is ampliative. However, it does not

seem to be a genuine case of abduction, since the empirical adequacy of a theory H (i.e., the truth of *all* observational consequences of H) does not *explain* the truth of the *so far* observed consequences of H. Rather, such an inference follows the pattern of enumerative induction, which involves the idea that the deductive consequences of theories are confirmed, and does not satisfy the abductive principle of Converse Entailment CE. More precisely, this inference follows a modification of the principle of Special Consequence SC in Section 4:

(SC') If evidence E as-if-confirms theory H, and H logically entails an observational statement E', then E confirms E'.[15]

Our counter-arguments against van Fraassen seem to indicate that his only interesting objection to abduction concerns cases where the prior probability P(H) is zero. But it is utterly implausible to make this assumption about all possible hypotheses that may occur to us in our ordinary lives or in science. Hence, he seems to be a kind of selective sceptic after all. In this respect van Fraassen has further developed this position by his non-standard Bayesian theory.

If H is a scientific hypothesis, the choice P(H) = 0 means an a priori commitment to scepticism about H, and this commitment can be claimed to be dogmatic, since updating this probability by Bayesian conditionalization always leads to a posterior probability of zero. Van Fraassen has suggested that *agnosticism* or "negative suspension" concerning a hypothesis H can be represented by an interval of probabilities [0,p) which includes 0 as its left point. In other words, agnosticism is "vagueness over an interval that includes 0" (see Hajek, 1998; Monton, 1998). This definition has the consequence that after updating by conditionalization, the posterior probability of H is still of the form [0,q), but q may be larger than p. In any case, this interval still includes the point 0, and in this sense negative suspension is a position that cannot cured by empirical learning. Van Fraassen (1998) also defends the view that such a negative suspension is an "eminently suitable epistemic attitude toward science". In particular, such a probability interval should be assigned to the theoretical claim about the existence of quarks.

In my view, one may question the rationality of a probability assignment which *a priori* allows one to rule out the existence of quarks, and still does this after the explanation of other particles in terms of quark theory. Even if our prior beliefs are vague, the interval (0,p) should not include 0 as its left point. But then all the results about Bayesian learning will be valid, and quark theory may be confirmed or refuted by evidence. This should be the position of a "good empiricist". What is more, such an empiricism is compatible with critical scientific realism: we have seen in Section 7 that even in cases where P(H) = 0 or P(E/H) = 0, it is possible to give an account of the cognitive significance of theory H in terms of its estimated truthlikeness: a scientific hypothesis may be known to be false, but it may still satisfy our desiderata by being closer to the truth than its rivals. The best method for searching for such good theories is to abductively

propose them as explanations of surprising facts (i.e., our earlier scientific results or conflicts) and then to assess them by empirical tests.

NOTES

1. See Peirce (1931-35) and (1958) [*CP*], (1982) [*W*], (1992a), (1992b) [*EP*1], (1998) [*EP*2]; Niiniluoto (1999b).
2. Cf. the distinction between fact-abduction and law-abduction in Schurz (2003). For similar ideas in AI research, see Josephson and Josephson (1994); Flach and Kakas (2000).
3. When Hempel in 1958 used the term 'retrodiction', he referred to Gilbert Ryle and W.H. Walsh. See Hempel (1965, p. 173).
4. Cf. the discussion of J.S. Mill's notion of analogy in Niiniluoto (1988).
5. In those cases where the probabilistic "validity" of the abductive inference (15) is high, Fumerton (1980) has suggested that it could be reduced to a probabilistic argument with the premise 'The proportion q of the Gs are F'.
6. Unlike Peirce, Whewell called such inference "induction" (see Niiniluoto, 1995).
7. Cf. the conception of hypothetico-inductive inference in Niiniluoto and Tuomela (1973).
8. It is interesting to observe that Bernard Bolzano's *Wissenschaftslehre* (1837) includes a chapter on the "art of discovery" (Erfindungskunst). In the English translation, the title of the chapter is "Heuretic". It discusses, among other things, "the discovery of causes for given effects" (see Bolzano, 1972).
9. See the criticism of this distinction in Howson and Urbach (1989, pp. 275-284).
10. Peirce was aware of this distinction, as we can see in *CP* 8.231-232.
11. Cf. Howson and Urbach (1989, pp. 270-275). For other treatments of the problem of old evidence, see Niiniluoto (1983) and Earman (1992).
12. Cf. Psillos (1996, 1999); Ladyman *et al.* (1997); Douven (1999); Okasha (2000).
13. The topic of this section is discussed in more detail in Niiniluoto (2003).
14. Another reply has been given by Theo Kuipers (1999, 2000), who applies his own comparative notion of truth-approximation. In his "downward" Success Theorem, Kuipers argues that higher truthlikeness explains greater empirical success, and in his "upward" Rule of Success he suggests that it is rational to favor a theory which has so far proven to be empirically more successful than its rivals. This gives a justification of *inference to the best theory*: "If a theory has so far proven to be the best one among the available theories, then conclude for the time being that it is the closest to the truth of the available theories."
15. Laudan (1984) also advocates anti-realism in the form that the truth of theories is irrelevant as an aim of science. Laudan (1990) offers "pragmatism" as a criticism of van Fraassen's "positivism", but his talk about "support" for, and "confirmation" of, a theory seems to mean something like support for the consequences of the theory (cf. Carnap's notion of instance-confirmation).

REFERENCES

Aliseda-Llera, A. *Seeking Explanations: Abduction in Logic, Philosophy of Science and Artificial Intelligence*. ILLC Dissertation Series, Amsterdam: University of Amsterdam 1997.

Bolzano, B., *Theory of Science*. Oxford: Blackwell 1972.

Boole, G. *The Laws of Thought*. New York: Dover 1958.

Douven, I. "Inference to the Best Explanation Made Coherent", in: *Philosophy of Science (Proceedings)* **66**, 1999, pp. S424-S435.

Earman, J. *Bayes or Bust? A Critical Examination of Bayesian Confirmation Theory*. Cambridge, MA: The MIT Press 1992.

Festa, R. "Bayesian Confirmation", in: Galavotti, M. and Pagnini, A. (eds.), *Experience, Reality, and Scientific Explanation*. Dordrecht: Kluwer 1999, pp. 55-87.

Fetzer, J.H. *Scientific Knowledge: Causation, Explanation, and Corroboration*. Dordrecht: D. Reidel 1981.

Flach, P.A. and Kakas, A.C. (eds.). *Abduction and Induction: Essays on their Relation and Integration*. Dordrecht: Kluwer 2000.

Fumerton, R.A. "Induction and Reasoning to the Best Explanation", in: *Philosophy of Science* **47**, 1980, pp. 589-600.

Hajek, P. "Agnosticism meets Bayesianism", in: *Analysis* **58**, 1998, pp. 199-206.

Hanson, N.R. "Is There a Logic of Discovery?", in: Feigl, H. and Maxwell, G. (eds.), *Current Issues in the Philosophy of Science*. New York: Holt, Rinehart, and Winston 1961, pp. 20-35.

Harman, G. "Inference to the Best Explanation", in: *The Philosophical Review* **74**, 1965, pp. 88-95.

Hempel, C.G. *Aspects of Scientific Explanation*. New York: The Free Press 1965.

Hintikka, J. "The Varieties of Information and Scientific Explanation", in: van Rootselaar, B. and Staal, J.F. (eds.), *Logic, Methodology, and Philosophy of Science III*. Amsterdam: North-Holland, 1968, pp. 151-171.

Hintikka, J. "The Place of C.S. Peirce in the History of Logical Theory", in: *Lingua Universalis vs. Calculus Ratiocinator. Selected Papers 2*. Dordrecht: Kluwer 1997, pp. 140-161.

Hintikka, J. "What is Abduction? The Fundamental Problem of Contemporary Epistemology", in: *Transactions of the Charles S. Peirce Society* **34**, 1998, pp. 503-533.

Hon, G. and Rakover, S.S. (eds.). *Explanation: Theoretical Approaches and Applications*. Dordrecht: Kluwer 2001.

Howson, C. and Urbach, P. *Scientific Reasoning: The Bayesian Approach*. La Salle: Open Court 1989.

Josephson, J. and Josephson, S. (eds.). *Abductive Inference*. Cambridge: Cambridge University Press 1994.

Kuipers, T. "Abduction aiming at Empirical Progress or even Truth Approximation leading to a Challenge for Computational Modelling", in: *Foundations of Science* **4**, 1999, pp. 307-323.

Kuipers, T. *From Instrumentalism to Constructive Realism: On Some Relations between Confirmation, Empirical Progress, and Truth Approximation*. Dordrecht: Kluwer 2000.

Ladyman, J., Douven, I., Horsten, L., and van Fraassen, B.C. "A Defence of van Fraassen's Critique of Abductive Reasoning: Reply to Psillos", in: *The Philosophical Quarterly* **47**, 1997, pp. 305-321.

Langley, P., Simon, H.A., Bradshaw, G.L. and Zytkow, J.M. *Scientific Discovery: Computational Explorations of the Creative Processes*. Cambridge, MA: The MIT Press 1987.

Laudan, L. *Science and Values: The Aims of Science and Their Role in Scientific Debate*. Berkeley: University of California Press 1984.

Laudan, L. *Science and Relativism*. Chicago: The University of Chicago Press 1990.

Levi, I. *Gambling with Truth*. New York: Alfred A. Knopf 1967.

Lipton, P. *Inference to the Best Explanation*. London: Routledge 1991.

Lipton, P. "What Good is an Explanation?", in: Hon and Rakover (2001a), pp. 43-59.

Lipton, P. "Is Explanation a Guide to Inference? A Reply to Wesley C. Salmon", in: Hon and Rakover (2001b), pp. 93-120.

Monton, B. "Bayesian Agnosticism and Constructive Empiricism", in: *Analysis* **58**, 1998, pp. 207-212.

Nesher, D. "Peircean Epistemology of Learning and the Function of Abduction as the Logic of Discovery", in: *Transactions of the Charles S. Peirce Society* **37**, 2001, pp. 23-57.

Niiniluoto, I. "Statistical Explanation Reconsidered", in: *Synthese* **48**, 1981, pp. 437-472.

Niiniluoto, I. "Novel Facts and Bayesianism", in: *The British Journal for the Philosophy of Science* **34**, 1983, pp. 375-379.

Niiniluoto, I. *Is Science Progressive?*. Dordrecht: D. Reidel 1984.

Niiniluoto, I. *Truthlikeness*. Dordrecht: D. Reidel 1987.

Niiniluoto, I. "Analogy and Similarity in Inductive Logic", in: Helman, D. (ed.), *Analogical Reasoning*. Dordrecht: D. Reidel, 1988, pp. 271-298.

Niiniluoto, I. "Peirce's Theory of Statistical Explanation", in: Moore, E.C. (ed.), *Charles S. Peirce and the Philosophy of Science*, Tuscaloosa and London: The University of Alabama Press, 1993, pp. 186-207.

Niiniluoto, I. "Descriptive and Inductive Simplicity", in: Salmon, W. and Wolters, G. (eds.), *Logic, Language, and the Structure of Scientific Theories*. Pittsburgh: University of Pittsburgh Press, 1994, pp. 147-170.

Niiniluoto, I. "Hintikka and Whewell on Aristotelian Induction", in: *Grazer Philosophische Studien* 49, 1995, pp. 49-61.

Niiniluoto, I. "Inductive Logic, Atomism, and Observational Error", in: Sintonen, M. (ed.), *Knowledge and Inquiry*. Amsterdam, Rodopi, 1997, pp. 117-131.

Niiniluoto, I. "Verisimilitude: The Third Period", in: *The British Journal for the Philosophy of Science* **49**, 1998, pp. 1-29.

Niiniluoto, I. *Critical Scientific Realism*. Oxford: Oxford University Press 1999a.

Niiniluoto, I. "Defending Abduction", in: *Philosophy of Science (Proceedings)* **66**, 1999b, pp. S436-S451.

Niiniluoto, I. "Abduction and Geometrical Analysis", in: Magnani, L., Nersessian, N. and Thagard, P. (eds.), *Model-Based Reasoning in Scientific Discovery*. New York: Kluwer and Plenum, 1999c, pp. 239-254.

Niiniluoto, I. "Hempel's Theory of Statistical Explanation", in: J.H. Fetzer (ed.), *Science, Explanation, and Rationality: The Philosophy of Carl G. Hempel*. Oxford: Oxford University Press, 2000, pp. 138-163.

Niiniluoto, I. "Abduction and Truthlikeness", in: R. Festa, A. Aliseda, and J. Peijnenburg (eds.), *Confirmation, Empirical Progress, and Truth-Approximation*, Rodopi, Amsterdam, forthcoming 2003.

Niiniluoto, I. and Tuomela, R. *Theoretical Concepts and Hypothetico-Inductive Inference*. Dordrecht: D. Reidel 1973.

Okasha, S. "Van Fraassen's Critique of Inference to the Best Explanation", in: *Studies in History and Philosophy of Science* **31**, 2000, pp. 691-710.

Peirce, C.S. *Collected Papers* 1-6, ed. by C. Hartshorne and P. Weiss, 7-8, ed. by A. Burks. Cambridge, MA: Harvard University Press 1931-35, 1958.

Peirce, C.S. *Writings of Charles S. Peirce: A Chronological Edition,* vols. 1-4, ed. by M.H. Fisch *et al.*, Bloomington: Indiana University Press 1982-.

Peirce, Charles S. *Reasoning and the Logic of Things: The Cambridge Conferences Lectures of 1898*, ed. by K.L. Ketner. Cambridge. MA: Harvard University Press 1992a.

Peirce, Charles S. *The Essential Peirce vol. 1 (1867-1893),* ed. by N. Houser and C. Kloesel, Bloomington: Indiana University Press 1992b.

Peirce, Charles S. *The Essential Peirce vol. 2 (1893-1913),* ed. by N. Houser and C. Kloesel, Bloomington: Indiana University Press 1998.

Popper, K.R. *The Logic of Scientific Discovery*. London: Hutchinson 1959.

Psillos, S. "On van Fraassen's Critique of Abductive Reasoning", in: *Philosophical Quarterly* **46**, 1996, pp. 31-47.

Psillos, S. *Scientific Realism: How Science Tracks Truth*. London: Routledge 1999.

Salmon, W. "Explanation and Confirmation: A Bayesian Critique of Inference to the Best Explanation", in: Hon and Rakover (2001a), pp. 61-91.

Salmon, W. "Reflections of a Bashful Bayesian: A Reply to Peter Lipton", in: Hon and Rakover (2001b), pp. 121-136.

Schurz, G. "Models of Abduction – From an Interrogative Viewpoint", in: *Synthese*, forthcoming 2003.

Smokler, H. "Conflicting Conceptions of Confirmation", in: *The Journal of Philosophy* **65**, 1968, pp. 300-312.

van Fraassen, B. *The Scientific Image*. Oxford: Oxford University Press 1980.

van Fraassen, B. *Laws and Symmetry*. Oxford: Oxford University Press 1989.

van Fraassen, B. "The Agnostic subtly Probabilified", in: *Analysis* **58**, 1998, pp. 212-220.

Department of Philosophy
P.O. Box 9
00014 University of Helsinki
Finland
ilkka.niiniluoto@helsinki.fi

STATHIS PSILLOS

INFERENCE TO THE BEST EXPLANATION AND BAYESIANISM*

Comments on Ilkka Niiniluoto's "Truth-seeking by Abduction"

1. INTRODUCTION

Niiniluoto (2003) has offered an incisive and comprehensive review of the recent debates about abduction. There is little on which I disagree with him. So, in this commentary, I shall try to cast some doubts to the attempts to render Inference to the Best Explanation (IBE) within a Bayesian framework.

Lately, there has been a lot of discussion about the place of IBE in Bayesian reasoning. Even Niiniluoto argues that "Bayesianism provides a framework for studying abduction and induction as forms of ampliative reasoning" (2003, 15). There is a tension, however, at the outset. Bayesian reasoning does *not* have rules of acceptance. On a strict Bayesian approach,[1] we can never detach the probability of the conclusion of a probabilistic argument, no matter how high this probability might be. So, strictly speaking, we are never licensed to accept a hypothesis on the basis of the evidence. All we are entitled to do, we are told by strict Bayesians, is a) to detach a conclusion about a probability, viz., to assert that the posterior probability of a hypothesis is thus and so; and b) to keep updating the posterior probability of a hypothesis, following Bayesian conditionalisation on fresh evidence. But IBE is typically seen as a rule of acceptance. In its least controversial form, IBE authorises the *acceptance* of a hypothesis H, on the basis that it is the best explanation of the evidence. Think of the standard IBE-based argument for the existence of middle-sized material objects. According to this, the best explanation of the systematic, orderly and coherent way we experience the world is that there are stable middle-sized material objects which cause our experiences. Presumably, those who endorse this argument do not just assert a conclusion about a probability; they assert a conclusion *simpliciter*. That is, their claim is not that the probability that material objects exist is high, but rather that it is reasonable to accept that they do exist. Hence, there is a tension between Bayesianism and standard renderings of IBE. This might make us wary of attempts to cast IBE in a Bayesian framework. But this is only the beginning of our worries.

Niiniluoto surveys a variety of recent results about the connection between IBE and Bayesian confirmation. They are all invariably instructive. But I want to challenge the *motivation* for attempting this. There are two questions to be asked.

F. Stadler (ed.), Induction and Deduction in the Sciences, 83–91.
© 2004 *Kluwer Academic Publishers.*

First, if we were to cast IBE within a Bayesian framework, could we do it? I do not doubt that this can be done (given the flexibility of the Bayesian framework). But I shall raise some worries about the ways that it can done. These worries will usher in the need the raise a *second* question, viz., should we want to cast IBE within a Bayesian framework? This question, I think, is more intriguing than the first.

2. IBE AND BAYESIAN KINEMATICS

The crux of IBE, no matter how it is formulated, is that explanatory considerations should inform (perhaps, determine) what it is reasonable to believe. Now, there are several ways to import explanatory considerations into a Bayesian scheme. There is a contentious one, due to Bas van Fraassen (1989). He claims that the right way to cast IBE within a Bayesian framework is to give *bonuses* to the posterior probabilities of hypotheses that are accepted as best explanations of the evidence. That is, *after* having fixed the posterior probability of a hypothesis in a Bayesian way, if this hypothesis is seen as the best explanation of the evidence, then it is entitled to a rise of its posterior probability. It's not hard to see that if one followed this way of updating one's degrees of belief, one would end up with Dutch books. In fact, this is precisely the strategy that van Fraassen himself follows in order to argue that, as a rule of updating degrees of belief, IBE is incoherent. But why should one take van Fraassen's suggestion seriously? As we have seen, the key recommendation of IBE is that explanatory considerations should inform (perhaps, determine) what we reasonably come to believe. So, if one were to cast IBE within a Bayesian framework, one should make sure that explanatory considerations are *part* of the Bayesian kinematics for the determination of the posterior probability of a theory, and not something that should be added on to confer bonus degrees of belief to the end product.

Given the Bayesian machinery, there are two ways in which explanatory considerations can be *part* of the Bayesian kinematics. They should either be reflected in the prior probability of a theory, relative to background knowledge, or (inclusively) in the likelihood of the theory. Niiniluoto shows what conditions should be satisfied vis-à-vis the priors and the likelihoods so that the best explanation is also the best confirmed theory (or that the better explanations receive the better confirmation). But better confirmation (even high confirmation) falls short of rightful acceptance. So, some of the excitement of IBE, as a rule of acceptance, is lost. But we have noted this already. An important further problem is this. Though priors and likelihoods *can* reflect explanatory judgements, it is clear that they fail to discriminate among competing hypotheses with the *same* priors and the *same* likelihoods. This problem is particularly acute for the case of likelihoods. The whole point of insisting on IBE is that it promises rationally to resolve observational ties. When two or more competing hypotheses entail the

very same evidence, then their likelihoods will be the same. Hence, likelihoods cannot resolve, at least some, (perhaps the most significant), observational ties.

But things get worse if we base our hopes on likelihoods. One thought, explored by Niiniluoto (2003, 22) might be to equate the best explanation with the hypothesis that enjoys the highest likelihood. But, as we are about to see, this is deeply problematic. In fact, as the so-called base-rate fallacy shows, likelihoods are relatively mute. If explanatory considerations enter the Bayesian story via likelihoods, then so much the worse for the explanatory considerations.

3. LIKELIHOODS AND THE BASE-RATE FALLACY

One way to introduce the base-rate fallacy is via the so-called Harvard Medical School test. Here are some details. A test for the presence of a disease has two outcomes, 'positive' and 'negative'. Let's call them *e* and *not-e*. Let a subject (Joan) take the test and let H be the hypothesis Joan has the disease. The test is highly reliable: it has zero *false-negative rate*. That is, the likelihood that the subject tested negative given that she *does* have the disease is zero (i.e., prob(not-e/H) = 0). Consequently, the *true-positive rate*, i.e., the likelihood of being tested positive given that she *does* have the disease is unity, (prob(e/H) =1). But the test also has a very small *false-positive rate*. That is, the likelihood that the subject is tested positive though she *doesn't* have the disease is, say, 5% (prob(e/not-H) = .05). Now, Joan takes the test and she tests positive. In the standard literature of the base-rate fallacy, given the above details, the following question is asked: what is the probability that Joan has the disease given that she tested positive? That is, what is the posterior probability prob(H/e)?

If we try to answer this question in a Bayesian framework, then it is clear that there is some crucial information missing: we are not given the incidence rate (base-rate) of the disease in the population. In other words, we are not given the prior probability of the hypothesis that the subject has the disease before she takes the test, i.e., prob(H). If this incidence rate is very low, e.g., if only 1 in 1,000 in the population has the disease, then it can be easily shown that it is very *unlikely* that Joan has the disease given that she tested positive: prob(H/e) would be less than .02.

It is notorious that when this problem is posed to experimental subjects, they tend, with overwhelming majority, to answer that the probability that Joan has the disease given that she tested positive is very high – very close to 95%. The so-called base-rate fallacy is that experimental subjects who are given the problem above tend to neglect base-rate information (that is, they tend to neglect the prior probabilities),[2] *even when* they are given this information explicitly. Several conclusions have been drawn from it and the relevant literature is massive. (Characteristically, one of the conclusions is that ordinary people are not Bayesians; another one is that ordinary people do not reason rationally because they do not follow Bayes's rule.) Suppose we asked the experimental

subjects: what is the best explanation of the evidence? That is, given that Joan tested positive in a highly reliable test, what is the best explanation of this fact? Now, this is *not* a question about probabilities. It is more like a question about what it is reasonable to accept about this particular case. Hence, it would not, it seems, be unreasonable for them to argue that the best explanation of the evidence is that Joan has the disease. But let's leave all this to one side.[3] The point I want to focus on is not whether and in what sense the base-rate neglect is indeed a fallacy. My point is simply that the base-rate fallacy (no matter how one reads it) shows that it is incorrect just to equate the best explanation of the evidence with the hypothesis that has the highest likelihood. As we saw above, it turns out that, if we consider just the likelihood of a hypothesis, and if we think that this is the way to determine the best explanation, then there is no determinate answer to the question 'what is the best explanation of the evidence?'. A very small prior probability can dominate over high likelihood and lead to a very small posterior probability. Let me put the point in a more conspicuous way. If we try to cast IBE within a Bayesian framework by focusing on likelihoods (that is, by saying that the best explanation is the hypothesis with the highest likelihood), then intuitive judgements of best explanation and judgements of Bayesian confirmation may well come apart.

Surely more needs to be said at this stage and I cannot say it now. Let me just distinguish between two issues. One is: can we equate the best explanation with the hypothesis that has the highest likelihood? I have just shown that we cannot. The other issue is: can we accept a hypothesis as the best explanation of the evidence if its posterior probability is low? This is a tough question. But I do not want to give a straightforward negative answer to it. Of course, it's unlikely that Joan has the disease given that she tested positive, if we know that the base-rate of the disease in the population is very low. But unlikely things happen and we don't want to say that it's outright unreasonable to believe that an unlikely thing has happened (especially if this best explains the evidence).[4] In any case, we are not always (or most typically) in situations where we have definite probabilities available. Nor can reasonable belief be equated with highly probable belief. There is more (and perhaps less) to reasonable belief than high probability.[5]

The point about likelihoods I have just made generalises. Consider what is called the Bayes factor, i.e., the ratio of likelihoods prob(e/not-H)/prob(e/H). One might try to connect IBE with likelihoods as follows. If the Bayes factor is small, then H is a better explanation of the evidence *e* than not-H. For, the thought will be, there are two ways in which the Bayes factor can be minimised: either when *e* is very unlikely when H is false or when *e* is very likely when H is true. Now, we can see that a version of Bayes's theorem is this:

$$\text{prob(H/e)} = \text{prob(H)}/\ \text{prob(H)} + f\,\text{prob(not-H)},$$

where f is the Bayes factor, i.e., prob(e/not-H)/prob(e/H). Wouldn't we expect that the smaller the Bayes factor is, the greater is the posterior probability of the

hypothesis? Wouldn't we thereby find a way to accommodate IBE, via the Bayes factor, within Bayesianism? Well, as above, what really happens depends on the prior probability. The Bayes factor, on its own, tells almost nothing. I say 'almost nothing' because there is a case in which the prior probability of a hypothesis does not matter. This is when the Bayes factor is *zero*. Then, no matter what the prior prob(H) is, the posterior probability prob(H/e) is one. So, it's only when just one theory can explain the evidence (in the sense that the likelihood prob(e/not-H) is zero) that we can dispense with the priors. That's a significant result. But does it show that IBE is accommodated within Bayesianism? In a sense, it does. But this sense is not terribly exciting. If there was only one potential explanation, then it would be folly not to accept it. But this case is really exceptional. We are still left with the need to distinguish between grue and green!

The moral so far is double. On the one hand, likelihoods cannot capture the notion of a good (the best) explanation. Put in a different way, even if likelihoods could, to some extent, carry the weight of explanation, they couldn't carry *all* of this weight on their own. On the other hand, we need to take into account the prior probabilities before we draw safe conclusions about the degree of confirmation of a hypothesis.

4. EXPLANATION AND PRIOR PROBABILITIES

What then remains of the Bayesian kinematics as an (indispensable) entry point for explanatory judgements is the prior probabilities. Now, it is one thing to say that priors are informed by explanatory considerations and quite another thing to say that they *should* be so informed. No-one would doubt the former, but subjective Bayesianism is bound to deny the latter. So, we come to the crux. There are two ways to think of IBE within a Bayesian framework. The *first* pays only lip service to explanatory considerations. For all the work in degree-of-belief updating (or, as some Bayesians say, in maintaining internal coherence in an agent's belief-corpus) is done by the usual Bayesian techniques and, perhaps, by the much-adored appeal to the washing out of priors. It may be admitted that the original assignment of prior probabilities might be influenced by explanatory considerations but the latter are no less idiosyncratic (from the point of view of the subjective Bayesian) than specifying the priors by, say, consulting a soothsayer. If we think this way, IBE, in a loose sense, is rendered consistent with Bayesianism, but it loses much of its excitement. It just amounts to a *permission* to use explanatory considerations in fixing an initial distribution of prior probabilities.

The other way to think of IBE within a Bayesian framework is to take explanatory considerations to be a *normative* constraint on the specification of priors. This is the way I would favour, if I were to endorse the Bayesianisation of IBE. It would be an exciting way to bayesianise IBE. (Better put, it would be an

exciting way to explanationise Bayesianism – forgive me the bad English words.) For it would capture the idea that explanatory considerations should be a rational constraint on inference. We might still be short of acceptance, since all we end up with is a degree of belief (no matter how high), but it would, at least, be a degree of *rational* belief. This move would also show how the resolution of observational ties is not an idiosyncratic matter. For some theories would command a higher initial rational degree of belief than others and this would be reflected, via Bayesian kinematics, in their posterior probability.[6]

But don't we all know that the story I have just outlined is, to say the least, *extremely* contentious? It would call for an objectivisation of Bayesianism and this is something that we, presumably, know it cannot be done. Whence do the explanatory virtues get their supposed rational force? And how are they connected with truth? I think these are serious worries. I am not sure they are compelling. For instance, I think there can be an a posteriori argument to the effect that theories with the explanatory virtues are more likely to be true than others (cf. my 1999, 171-6). And there is also an argument to the effect that judgements of prior probabilities should aim to improve the coherence of our system of beliefs and that the explanatory virtues improve such coherence (cf. my 2002). But showing all this would be an uphill battle. It would call, to say the least, for a radical departure from the standard Bayesian criteria of rationality and belief-revision. So, the project of accommodating IBE within Bayesianism would involve a radical rethinking of Bayesianism. And not many people are, nowadays, willing for such a radical rethinking.

5. A DILEMMA

The way I have described things leads us to a dilemma. *Either* accommodate (relatively easily) IBE within Bayesianism but lose the excitement and most of the putative force of IBE *or* accommodate an interesting version of IBE but radically modify Bayesianism. I guess we all agree that Bayesianism is the best theory of confirmation available. But at least some of us are unwilling to think that Bayesianism is the final word on the matter, since we think that there is more to rationality (and to scientific method) than Bayesians allow. Those of us who are friends of IBE might then have to reject the foregoing dilemma altogether. This would bring us back to the second question I raised in section 1, and which I took to be the more interesting one: *should we want to cast IBE within a Bayesian framework?*

I cannot start answering this question in this paper. I hope to have sketched why there are reasons to take it seriously. I will conclude with a note on how a *negative* answer to it can be motivated. IBE is supposed to be an *ampliative* method of reasoning. It is supposed to deliver informative hypotheses and theories, viz., hypotheses and theories which exceed in content the observations, data, experimental results etc. which prompt them. This content-increasing as-

pect of IBE is indispensable, if science is seen, at least *prima facie*, as an activity that purports to extend our knowledge (and our understanding) beyond what is observed by means of the senses. Now, Bayesian reasoning is *not* ampliative. In fact, it does not have the resources to be ampliative. All is concerned with is maintaining synchronic consistency in a belief corpus and (for some Bayesians, at least) achieving diachronic consistency too. Some Bayesians, e.g., Colin Howson (2000), take probabilistic reasoning to be a mere extension of deductive reasoning, which does not beget any new factual content.

There might be two related objections to what I have just said. The first might be that Bayesian reasoning allows for ampliation, since this can be expressed in the choice of hypotheses over which prior probabilities are distributed. In other words, one can assign prior probabilities to ampliative hypotheses and then use Bayesian kinematics to specify their posterior probabilities. The second (related) objection may be found in what Niiniluoto says at some point, viz., that "the Bayesian model of inference helps to show how evidence may confirm hypotheses that are abductively introduced to explain them" (2003, 20). Here again, the idea is that abduction might suggest ampliative hypotheses, which are then confirmed in a Bayesian fashion. If we elaborate and combine the two objections in an obvious way, they imply the following: ampliative IBE and non-ampliative Bayesian reasoning might well work in tandem to specify the degree of confirmation of ampliative hypotheses.[7]

I too have toyed with this idea and still think that there is something to it. In an earlier piece I noted:

> although a hypothesis might be reasonably accepted as the most plausible hypothesis based on explanatory considerations (abduction), the *degree of confidence* in this hypothesis is tied to its degree of subsequent confirmation. The latter has an antecedent input, i.e., it depends on how good the hypothesis is (i.e., how thorough the search for other potential explanations was, how plausible a potential explanation is the one at hand etc.), but it also crucially depends on how well-confirmed the hypothesis becomes in light of further evidence. So, abduction can return likely hypotheses, but only insofar as it is seen as an integral part of the method of inquiry, whereby hypotheses are further evaluated and tested (2000, 67).

But we can also see the limitations of the idea under discussion. For what, in effect, is being conceded is that IBE (or abduction) operates *only* in the context of discovery, as a means to generate plausible ampliative hypotheses and to distil the best among them. Then, the best explanation is taken over to the context of justification, by being embedded in a framework of Bayesian confirmation, which determines its credibility. I think the friends of IBE have taken IBE to be both ampliative *and* warrant-conferring at the same time. It is supposed to be an ampliative method that confers warrant to the best explanation of the evidence. So, if we are concerned with giving a precise degree of confirmation to the best explanation, or if we subject it to further testing, then we can indeed embed it in a Bayesian framework. But something would have gone amiss if we thought that

the best explanation was *not* reasonably acceptable before it was subjected to Bayesian confirmation. To see how this reasonable acceptance could be analysed would lead us beyond the confines of this commentary. But I have started working out the details elsewhere (cf. 2002). For, I think we can profitably cast the issue of the warrant of IBE within the theories of justification which connect justification with the absence of defeaters, theories which were made popular by John Pollock (1986).

NOTES

* Many thanks to Ilkka Niiniluoto, Peter Lipton, Maria-Carla Galavotti and the participants of the Workshop "Induction and Deduction in the Sciences" for many useful comments on an earlier draft.

1. Niiniluoto (personal communication) has rightly pointed out that there are two big strands within Bayesianism. One of them (Levi, Hintikka) promotes the idea of inductive acceptance rules, and hence, advocates ampliative inferences. The other branch (Carnap, Jeffrey, Howson) rejects acceptance rules and considers only changes of probabilities. It is this latter strand of Bayesianism that I take issue with at this point.
2. It is very debateable that we should equate the base-rate with the prior probability. But nothing hangs on this in the use I make of the base-rate fallacy.
3. There is a lot of recent re-evaluation of the base-rate fallacy (cf. Koehler 1996). One point that seems worth making, though I am not sure I want to endorse it in full, is that the base-rate fallacy relates specifically to probabilistic reasoning, where reference classes are to be taken into account. We may or may not be good at seeing the need to take into account reference classes in order to draw conclusions about probabilities concerning individual cases. But it's not clear to me why we *should* take into account reference classes when we look for best explanations of the evidence. An explanation can be reasonable to accept (even true), even though it is unlikely. In a sense, what matters for the explanation of an individual event is not what the other members of the reference class we put it in do, but rather what the details of the individual case we are interested in are.
4. Consider the following. A Geiger-counter detects a certain type of particle by registering a click. The particles are very rare so that the probability that the counter clicks is very low. But suppose it does click. Is it not unreasonable to believe that it registered a particle, especially if it's highly reliable in doing this?
5. That there is more to reasonable belief than high probability is argued at length by Achinstein (2001, chapter 7). He takes high probability to be necessary for rational belief but he denies that it is sufficient. One of his explicit additional requirements is that there must be an explanatory connection between a hypothesis and its evidence.
6. There is an interesting idea in Niiniluoto's paper (2003, 21) that needs to be noted. One may call "systematic power" the explanatory and predictive power of a hypothesis relative to the total *initial* evidence *e*. Then, one may use this systematic power to determine the probability prob(H) of the hypothesis H. Prob(H) will be none other than the *posterior probability of H* relative to its ability to explain and predict the initial evidence *e*, that is relative to its systematic power. In this sense, it can be argued that the prior probability of a hypothesis does depend on its explanatory (that is, its systematic) power.
7. This line is pressed in a fresh and interesting way in Lipton's (2001). Lipton tries to show how a Bayesian and an Explanationist can be friends. In particular, he shows how explanatory considerations can help in the determination of the likelihood of a hypothesis, of its prior probability and of the relevant evidence. I think all this is fine. But it should be seen not as an attempt at peaceful co-existence, but rather as an attempt to render Bayesianism within the

Explanationist framework, and hence as an attempt to make Bayesianism an objectivist theory of confirmation.

REFERENCES

Achinstein Peter, *The Book of Evidence*, New York: Oxford University Press 2001.
Howson Colin, *Hume's Problem*, New York: Oxford University Press 2000.
Koehler Jonathan J., "The Base Rate Fallacy Reconsidered: Descriptive, Normative and Methodological Challenges", *Behavioural and Brain Sciences* 19, 1996, pp. 1-53
Lipton Peter, "Is Explanation a Guide to Inference", in: G. Hon & S. S. Rakover (Eds.), *Explanation: Theoretical Approaches and Applications*, Dordrecht: Kluwer 2001, pp. 93-120.
Niiniluoto Ilkka, "Truth-Seeking by Abduction" *this volume*.
Pollock John, *Contemporary Theories of Knowledge*, Rowan & Littlefield 1986.
Psillos Stathis, *Scientific Realism: How Science Tracks Truth*, London: Routledge 1999.
Psillos Stathis, "Abduction: Between Conceptual Richness and Computational
Complexity" in: A. K. Kakas and P. Flach (Eds.), *Abduction and Induction: Essays in their Relation and Integration*, (Applied Logic Series, Vol. 18), Dordrecht: Kluwer 2000, pp. 59-74.
Psillos Stathis, "Simply the Best: A Case for Abduction" in: Fariba Sadri & Anthony Kakas (Eds.), *Computational Logic: From Logic Programming into the Future*, LNAI 2408, Berlin-Heidelberg: Springer-Verlag 2002, pp. 605-25.
van Fraassen Bas, *Laws and Symmetry*, Oxford: Clarendon Press 1989.

Department of Philosophy and History of Science
University of Athens
Panepistimioupolis (University Campus)
Athens 15771
Greece
psillos@phs.uoa.gr

JOKE MEHEUS[†]

ADAPTIVE LOGICS AND THE INTEGRATION OF INDUCTION AND DEDUCTION[*]

1. INTRODUCTION

The aim of this paper is twofold. First, I want to argue that the distinction between induction and deduction is less clear-cut than traditionally assumed, and that, moreover, most reasoning processes in the sciences involve an integration of inductive and deductive steps. Next, I want to show how so-called adaptive logics may lead to a better understanding of this integrated use of induction and deduction.

Here and below, I use the term "induction" in a broad sense to refer to any inference in which the conclusion does not necessarily follow from the premises, but is only (sufficiently) likely in view of them. As usual, I call an inference "deductive" if it is truth-preserving. Note that the term "premises" is ambiguous. It may either refer to the statements on which an individual inference is based or to the statements that form the starting point of some reasoning process (a series of inferences).[1] As the distinction will be important in this paper, I shall always make sure that the ambiguity is resolved by the context.

An important difference between deductive and inductive inferences is that the latter give rise to *non-monotonic* reasoning processes: a conclusion inferred against the background of some set of premises may be withdrawn when this set is extended. As an example one may think of generalizations ("All ravens are black") that are, at some moment in time, inferred from a set of positive instances, but that are withdrawn when this set is, at a later moment in time, extended with some negative instance. An immediate consequence of this is that inductive inferences are *conditional* in nature: their conclusion is accepted on the condition that some other conclusion is not inferred from the background premises (for instance, a negative instance). This is not the case for deductive inferences: the acceptance of their conclusion is only dependent on the acceptance of their premises – if the latter are accepted, then so is the former.

When the distinction is characterized in this way, it seems natural to assume that there exists some general criterion to classify different inferences (and, by extension, different reasoning processes) into inductive and deductive. This assumption is further supported by the situation in formal logic. Although all logicians agree that formal logic can successfully be applied to inferences tradi-

F. Stadler (ed.), Induction and Deduction in the Sciences, 93–120.
© 2004 *Kluwer Academic Publishers.*

tionally considered as deductive (such as inferences of the form $A \supset B, A / B$ and $A \supset B, \neg B / \neg A$), most of them are highly sceptical that it can also be applied to inference patterns traditionally classified as inductive (such as inferences of the form $(\forall \alpha)(A(\alpha) \supset B(\alpha)), B(\beta) / A(\beta)$). The reason for this scepticism is simple. In the case of deductive inferences, soundness can adequately be captured by means of some formal rule. This seems to be impossible, however, in the case of inductive inferences: the only sensible formal rule behind them seems inevitably to lead to a set of unsound and even inconsistent conclusions. Hence, in as far as inductive inferences are sound, it seems that they must rely on a hidden non-formal reasoning.

However, when turning to actual cases of reasoning, the assumption that induction and deduction can neatly be separated does not seem to be very helpful. For one thing, it seems difficult to find examples of reasoning processes that, on the above characterization, are purely deductive. In most reasoning processes, deductive and inductive steps are constantly interchanged. Moreover, in a majority of these cases, the two kinds of steps are not simply alternated, but depend on one another: deductive inferences are applied to conclusions of inductive inferences and *vice versa*. Hence, it seems impossible to understand the former while leaving out the latter.

To most readers, this will seem rather evident. They will connect it to the fact that, both in the sciences and in everyday situations, we usually have to reason under conditions of incomplete knowledge, and hence, are forced to apply inference patterns traditionally considered as inductive (such as generalizations, abduction, analogies, ...). Indeed, almost by definition, it seems that pure deduction is only possible in cases of complete knowledge – deductive steps reveal nothing that is not already implicitly present in the premises.

There is, however, something else. When carefully looking at examples of actual reasoning, it turns out that the distinction between induction and deduction is less clear-cut than traditionally assumed: inferences that in one context clearly classify as deductive may in a different context appear to be inductive (because they are conditional and give rise to non-monotonic reasoning patterns). There are, for instance, examples of reasoning processes in which no other inferences occur than those validated by Classical Logic (henceforth, **CL**), but in which nevertheless some of the inferences, on the above definitions, should be classified as inductive.[2]

In view of observations like these, many philosophers of science, since Kuhn, concluded that scientific reasoning 'goes beyond logic', and hence, that "rationality" should be disconnected from "logic". In line with this, they no longer relied on logic to explicate the mechanisms underlying scientific reasoning, but instead turned to methodological analyses, to results from the cognitive sciences, or to a combination of historical and cognitive methods.[3]

This turn away from logic has been extremely important: it made it possible to study the enormous richness of reasoning patterns that, indeed, do not fit within the traditional framework of formal logic. Today, however, the tools are

available to study these reasoning patterns also from a logical point of view. The importance of this is not that any of the existing approaches can or should be replaced by one that is logic-based, but rather that logic can contribute to obtain a more complete analysis of scientific reasoning. For instance, although cognitive approaches led to a better understanding of many of the mechanisms behind scientific reasoning that are indeed extra-logical (such as the retrieval of relevant information), they are not well-suited for the appraisal of reasoning processes (in the sense that they do not enable one to distinguish between sound and unsound inferences). The latter is typically a problem that asks for a logic-based approach.

In this paper, I shall show how scientific reasoning processes can be explicated in terms of adaptive logics. As we shall see below, adaptive logics provide a unified framework for the formal study of reasoning processes that exhibit an external and/or an internal dynamics – I call the dynamics external, if it is caused by the addition of new premises (non-monotonicity) and internal, if it is caused by the mere analysis of the premises.[4]

At this moment, a whole family of adaptive logics has been formulated. Some of these validate inference rules traditionally considered as inductive, others only validate the inference rules of **CL**. However, none of them seems to fit into the traditional induction-deduction dichotomy. The reason is that all adaptive logics integrate the two kinds of inferences. Precisely because of this, they are well-suited to understand the combined use of deductive and inductive inferences as we know it from everyday reasoning and the sciences.

Adaptive logics are not the only tools currently available for the formal study of non-monotonic reasoning processes. As compared to other tools, however, adaptive logics have the enormous advantage that they allow for a *realistic* explication of actual reasoning processes – one that, for instance, does not presuppose logical omniscience. This is related to a peculiar property of the proof theory of adaptive logics: at any stage of a proof, it is allowed that inferences are made on the basis of the best insights in the premises *at that stage of the proof*. As a consequence, the proof theory of adaptive logics is dynamic – conclusions derived at some stage in the proof may at a later stage be withdrawn. In view of this dynamics adaptive logics are characterized by two notions of derivability: derivability at a stage and final derivability. For most adaptive logics currently available, the latter notion has been proven sound and complete with respect to a proper semantics.

In this paper, I shall proceed as follows. In Section 2, I present some examples from the sciences that illustrate the combined use of inductive and deductive steps. The main example in this section concerns Clausius' derivation of Carnot's theorem from an inconsistent set of premises, and illustrates how inferences traditionally considered as deductive can nevertheless give rise to a non-monotonic reasoning process. In Section 3, I discuss which challenges reasoning processes like these pose for a logical study. The next two sections are devoted to adaptive logics. After briefly discussing the basic ideas behind them (Section

4), I offer a generic formulation of their dynamic proof theory (Section 5). The latter enables me to present in a concise way adaptive logics for three main types of dynamic reasoning: an inconsistency-adaptive logic, a logic for inductive generalizations, and a logic for abduction (Section 6). The first of these is suited to understand the example from Section 2, the second enables one to infer empirical generalizations as well as predictions from a set of empirical data and the third to infer explanatory hypotheses and predictions from a set of singular statements (the explananda) and a background theory. Some conclusions are presented in Section 7.

I end this section with some remarks on the use of the term "logic". In this paper, I use this term in a broad sense to refer to any system that enables one, for a specific application context, to distinguish between sound and unsound reasoning, and that, moreover, is characterized, in a formally decent way,[5] by a semantics and a proof theory. Adaptive logics meet all these requirements. However, I shall not quibble if some readers prefer to reserve the term "logic" for systems that satisfy some additional requirements, and hence, want to name adaptive logics otherwise. The essential point is that adaptive logics enable one to explicate, in a formally exact and nevertheless realistic way, people's actual reasoning processes. How they are called is rather irrelevant to this.

2. SOME EXAMPLES FROM THE SCIENCES

I mentioned already in the previous section that, under conditions of incomplete knowledge, deductive steps are continuously combined with inductive ones, and that the resulting reasoning process is non-monotonic. When seeking an explanation for a novel fact, for instance, explanatory hypotheses are often formulated on the basis of abductive steps. These hypotheses, however, are withdrawn when their negation is (deductively) derived from the background theory.

An example of such an explanation-seeking process concerns the discovery of Uranus. When Herschel in 1781 noted an unexpectedly large object in the quartile near Zeta Tauri, he first formulated the hypothesis that the body at issue was a comet. This was inferred from, on the one hand, the fact that the object appeared especially large and that it moved among the stars, and, on the other hand, the theoretical assumption that comets have both these properties. When the hypothesis was contradicted by observational data concerning the object's orbit, it was withdrawn and a new hypothesis was formulated (namely, that the body was a planet). The latter too was obtained by an abductive step.[6] Also the reasoning process that led to the discovery of oxygen incorporated several abductive inferences, many of which were withdrawn in view of other information available.[7]

Other examples that fit into this category are analogical reasoning processes. Well-known examples include Bohr's design of his model of the atom (on the basis of an analogy with the solar system), and Carnot's formulation of his theory of the steam engine (on the basis of an analogy with hydraulic engines). Also

here, inductive conclusions may be withdrawn (for instance, when it turns out that one of the properties of the source domain is not applicable to the target domain).

All examples thus far mentioned involve *ampliative* inference patterns – I say that an inference rule is ampliative if it is not validated by **CL**, and, moreover, can only be added to **CL** at the expense of obtaining a system that leads to unsound conclusions. This, however, is not a necessary characteristic of a non-monotonic reasoning process. My favourite example of a reasoning process that involves no other inferences than those validated by **CL**, but that is nevertheless dynamic, is Clausius' derivation of Carnot's Theorem. As this kind of example is less well-known than the above, I shall discuss it in some more detail.

One of the central contributions to early thermodynamics was Carnot's derivation of, what was later called, Carnot's Theorem. According to this Theorem, no engine can be more efficient than a reversible one. Carnot's derivation was published in 1824,[8] and was based on the calorique theory of heat. As Carnot's Theorem proved extremely successful, it was soon after its publication considered as one of the most fundamental results in the domain. Around 1840, however, it became clear that the calorique theory of heat was contradicted by some reliable experimental results concerning the conversion of work into heat. These results, moreover, supported the rival view that heat is not a substance, but some kind of 'force' that can be converted into other types of 'forces'. Some of the leading scholars in the domain, like Kelvin, concluded from this that one would have to restart from scratch, but that, in the meantime, it was better to stick to Carnot's theory. Others, like Clausius, tried to reason from the two incompatible approaches in the hope to arrive at a consistent alternative.[9]

As may be expected, one of the main challenges concerned the problem to find a new derivation for Carnot's Theorem. In 1850, Clausius presented such a new derivation that, although the inconsistencies were not yet resolved, was accepted by the scientific community.[10] What is even more interesting, however, is that, according to Clausius' own account, he designed two proofs, the first of which he rejected.[11]

As I described elsewhere,[12] both proofs are very similarly to one another – they start from the same set of premises, and are both based on *Reductio ad Absurdum*. Thus, where C stands for Carnot's Theorem, both $\neg C \supset D$ and $\neg C \supset \neg D$ are derived in the first proof, and $\neg C \supset E$ and $\neg C \supset \neg E$ in the second.[13] Next, *Reductio ad Absurdum* is applied in both proofs to obtain C. Given this structural similarity, it seems highly surprising that Clausius rejected the first proof, but accepted the second one. The reason for this becomes clear, however, if one takes a closer look at the proofs, and moreover takes into account that the set of premises at issue was inconsistent.[14]

A crucial difference between the two proofs is that, in the case of the first proof, D behaves inconsistently with respect to the background premises: even if one proceeds in a sensible way (that is, while avoiding *Ex Falso Quodlibet*) both D and $\neg D$ are derivable from them. But, if both D and $\neg D$ are true,

then so are $\neg C \supset D$ and $\neg C \supset \neg D$, even if C is false. Hence, in this case, it would indeed be unsound reasoning to apply *Reductio ad Absurdum* and to conclude that C holds true. In the case of the second proof, however, E behaves consistently with respect to the background premises: if one proceeds sensibly, $E \wedge \neg E$ cannot be derived from them. Now, unless one is forced to accept otherwise, it seems reasonable to assume that $E \wedge \neg E$ is false. But, if $E \wedge \neg E$ is false, then $\neg C \supset E$ and $\neg C \supset \neg E$ can only be true if $\neg C$ is false. So, in this case, it seems to make perfectly good sense to apply *Reductio ad Absurdum* and to conclude that C holds true.

Evidently, I do not claim here that Clausius was aware of the kind of logic he was using. I do claim, however, that he seemingly had some very good logical intuitions. As soon as he discovered that D behaves inconsistently with respect to the background premises, there were indeed good logical grounds to reject the derivation and to start anew. In the next sections, we shall see how these intuitions can be made more precise.

3. THE CHALLENGES

In this section, I shall argue that the examples presented in the previous section confront logicians with a double challenge. But first, I need to analyse the examples a bit further, and show what they have in common from a formal point of view. Let us first turn to the example of Clausius.

Readers familiar with paraconsistent logics[15] will immediately have recognized my argument for the unsoundness of Clausius' first proof. Within paraconsistent logic, it is quite commonly agreed upon that, as soon as one allows for the (non-trivial) acceptance of inconsistencies, *Reductio ad Absurdum* is no longer truth-preserving.[16] However, it is not the case that Clausius simply invalidated this rule of inference. Rather, he invalidated some of its *applications*, while validating others. So, some inference rule, which is in the context at issue not truth-preserving, is nevertheless applied, but *subject to some condition*. If this condition is not (or no longer) satisfied, then the application is invalidated. Characterized in this way, Clausius' reasoning process becomes very similar to one in which, for instance, abduction is applied or in which inductive generalizations are derived. Also here, the conditional application is allowed of one or more inference rules that are not truth-preserving.

This brings us to the first challenge posed by this type of reasoning processes: to obtain a formal grip on the conditional validation of inference rules that do not preserve truth.

An obvious answer to this challenge seems to be that we include the conditions in the formulation of the inference rules. Given the nature of the conditions, this requires that we refer to the set of background premises. In the example of Clausius, for instance, $\neg A$ is derived from $A \supset B$ and $A \supset \neg B$, on the condition that $B \wedge \neg B$ is not derivable from the premises against the background of

which the inference is applied. At first sight, this reference to the background premises may seem objectionable. However, the formal character of a logic does not derive from the fact that there is a formal link between the premises of the application of some rule and its conclusion, but from the fact that there is a formal link between (a subset of) the background premises and the last step in a proof. In the case of monotonic logics, the formal character of the logic warrants the formal character of the rules. In the case of non-monotonic logics, it does not. Here, as we have seen, the soundness of an application of a rule may depend on the background premises.

Given this convention on the formal character of a logic, the reconstruction of the logic behind Clausius' reasoning seems to be extremely simple. It seems to be obtained by replacing, in an appropriate manner, every inference rule of **CL** that, when applied to an inconsistent set of premises, is not truth-preserving. Thus, where **PL** stands for some paraconsistent fragment of **CL**, an adequate replacement for *Reductio ad Absurdum* seems to be:[17]

(1) If $\Gamma \vdash_{\mathbf{PL}} A \supset B$, $\Gamma \vdash_{\mathbf{PL}} A \supset \neg B$, and $\Gamma \nvdash_{\mathbf{PL}} B \wedge \neg B$, then $\neg A$ may be added to any proof of Γ in which both $A \supset B$ and $A \supset \neg B$ occur.

In the case of the other examples, it seems that the conditions may be somewhat more elaborated, but that the mechanism is the same.

There is, however, a price to be paid when this road is followed. At the predicative level, the application of (1) is undecidable. Indeed, as all interesting paraconsistent fragments of **CL** are undecidable at the predicative level, it may be impossible to establish, for some Γ and some B, that $\Gamma \nvdash_{\mathbf{PL}} B \wedge \neg B$, and hence, to establish that (1) can be applied.

A different way to put this is that the consequence relation behind Clausius' reasoning lacks a positive test: even if some formula is derivable according to it, there need not exist a finite construction that establishes this.[18] This lack of a positive test not only holds true for the example of Clausius, but for all forms of non-monotonic reasoning. The reason is simple: if a reasoning process is non-monotonic, then some of its inferences are conditional, and the condition always refers to the fact that some specific formula is *not* derived.

The second challenge is related to this: the lack of a positive test should not make the applicability of the inference rules undecidable.

Basically, there seem to be two ways to meet this challenge. The first is that one only considers decidable theories. For such theories, one may be sure that, if some formula is not derivable, one will be able to establish this by a finite construction. The second is that one allows that inferences are made, not on the basis of absolute warrants, but on the basis of one's best insights in the premises. When this second option is followed, the resulting reasoning processes not only exhibit an external form of dynamics, but also an internal one.

There are several arguments in favour of this second option. The first is that unwanted restrictions are avoided: *any* first-order theory can be dealt with. A second argument is that the conclusions arrived at in a non-monotonic reasoning process are defeasible anyway. Whether the withdrawal is caused by an external factor or an internal one does not seem to be essential. The third, and most important argument is that, even for decidable fragments, it is often unrealistic to require absolute warrants. Reasoners (whether human or artificial) are not logically omniscient: discovering the more interesting consequences of one's theory requires time and effort. So, even if a decision method is available, reasoners may lack the resources to perform an exhaustive search, and hence, may be forced to act on their present best insights. This is also illustrated by the example of Clausius. If his own account is correct, he completed his first proof, and only later discovered that some of the statements involved behaved inconsistently with respect to the premises.

As hinted at already in the first section, adaptive logics follow the second option. This has the advantage that, even for undecidable fragments, they enable one to come to justified conclusions. These conclusions are tentative and may later be rejected, but they constitute, given one's insight in the premises at that moment, the best possible estimate of which conclusions are derivable from the premises.

4. ADAPTIVE LOGICS

The first adaptive logic was designed around 1980 by Diderik Batens, and was meant to handle in a sensible and realistic way inconsistent sets of premises.[19] This logic was followed by several other inconsistency-adaptive systems, and several inconsistency-handling mechanisms that proceed in terms of maximal consistent subsets were reconstructed in terms of adaptive logics and even strengthened.[20] Later the idea of an adaptive logic was generalized to other forms of logical abnormalities, such as negation-incompleteness and ambiguity and to other types of premises, such as questions.

Until a few years ago, all adaptive logics were *corrective* systems (they can handle logically abnormal theories – for instance, inconsistent ones). It was shown, however, that exactly the same kind of formal techniques can be used to explicate *ampliative* forms of reasoning.[21] Meanwhile, (formal) results are available on adaptive logics for compatibility, pragmatic truth, diagnostic reasoning, inductive generalizations, abduction and inference to the best explanation, and the analysis of metaphors.[22]

What all these logics have in common is that, as their name implies, they adapt themselves to the specific set of premises to which they are applied. This means that, at least for some of the inference rules, it depends on properties of the premise set whether their application is validated or not. Importantly, this does not require any intervention of the user – the logics really adapt *themselves*

to the premises. Moreover, as we shall see below, this adaptation proceeds in a *contextual way* – even if the application of an inference rule is invalidated with respect to some 'parts' of the premises, it may be validated with respect to others.[23]

This contextual validation of inference rules is related to the application context for which adaptive logics are meant, namely reasoning processes that exhibit an external and/or internal dynamics. A typical characteristic of such processes is that a specific set of presuppositions is maintained 'as much as possible', this is, *unless and until* they are explicitly violated. For instance, it is not possible to reason sensibly from an inconsistent set of statements while presupposing that $A \wedge \neg A$ is false, for any possible A. Still, one may maintain this presupposition for all A that have not explicitly been shown to behave inconsistently. If this presupposition is violated, because some B has been found to behave inconsistently, then previously derived conclusions may be withdrawn.

Until quite recently, the existing formal systems were not fit for reasoning processes like these. In most logics, the violation of presuppositions is sanctioned with triviality. **CL**, for instance, presupposes consistency as well as negation-completeness, and turns any theory that violates these presuppositions into the trivial one.

In order to deal with theories that violate one or more **CL**-presuppositions, a whole variety of non-classical logics was designed. Most of these are obtained by simply dropping some **CL**-presuppositions and by restricting the inference rules accordingly. Examples are Jaśkowski's **D2**, da Costa's **C**-systems, and Priest's **LP**. All of these drop the consistency presupposition, and restrict the rules of inference in such a way that *Ex Falso Quodlibet* is invalidated. Also, these logics are inadequate for dynamic reasoning processes. Not only do they fail to capture the dynamics involved, they are too poor to make sense of examples like that of Clausius.[24]

What makes adaptive logics suitable for dynamic reasoning processes is that they adapt themselves to specific violations of presuppositions. Where a presupposition is violated, the rules of inference are restricted in order to avoid triviality. However, where this is not the case, the rules can be applied in their full strength. As a consequence, and as mentioned a few paragraphs ago, the rules of inference become context-dependent: whether they are applicable or not depends on properties of the premises to which they are applied.

Some adaptive logics are adequate for situations in which a preference ranking is defined over the background premises. If they are, they are called *prioritized*. In the rest of this paper, I shall only consider flat (that is, non-prioritized) adaptive logics.[25] Formally, a flat adaptive logic **AL** may be characterized in terms of three elements: a lower limit logic, a set of abnormalities and an adaptive strategy.[26]

The lower limit logic (**LLL**) is monotonic and characterized by $Cn_{\mathrm{LLL}}(\Gamma) = \bigcap\{A \mid \Gamma \cup \Delta \vdash_{\mathrm{AL}} A ; \varnothing \subseteq \Delta \subseteq \mathcal{W}\}$ in which $\mathcal{W}$ is the set of all closed

formulas of the language. From a proof theoretic point of view, the lower limit logic defines the rules of inference that hold unconditionally. From a semantic point of view, it provides the models from which the adaptive models of Γ are selected. (As we shall see below, all adaptive models of Γ are lower limit models of Γ, but not conversely.) In the case of an inconsistency-adaptive logic, for instance, the lower limit logic is some monotonic paraconsistent logic. In the case of the logic of abduction discussed below, the lower limit logic is **CL**.

The set of abnormalities Ω consists of the formulas that are presupposed to be false, *unless and until* proven otherwise. For instance, in handling inconsistencies, the set of abnormalities is some set of formulas of the form $\exists(A \wedge \neg A)$ – the existential closure of $A \wedge \neg A$. In the case of the logic for inductive generalizations that we shall see below, the set Ω consists of formulas of the form $\exists A \wedge \exists \neg A$. Note that the term "abnormality" does not refer to the purported standard of reasoning, say **CL**. It refers to properties of the application context – to presuppositions that are considered desirable, but that may be overruled.

An important constraint for the set of abnormalities is that extending the lower limit logic with the requirement that no abnormality is logically possible should result in a monotonic logic. This logic is called the upper limit logic (**ULL**). Semantically, the upper limit logic is characterized by the lower limit logic models that verify no abnormality. Thus, the upper limit logic defines the 'normal situation' – the one in which all formulas behave normally. As a consequence, it turns any abnormal theory into the trivial one.

All adaptive logics developed by the 'Ghent group' (Diderik Batens and associates) take **CL** as their standard of reasoning. In line with this, the corrective ones have **CL** as their upper limit logic and a *fragment* of this as their lower limit logic, and the ampliative ones have **CL** as their lower limit logic and an *extension* of this as their upper limit logic.

At this point, it may seem that an adaptive logic is obtained by presupposing that all abnormalities (members of Ω) are false, except for those that, by the lower limit logic, follow from the premises. There is, however, a small complication. This is related to the fact that, for most lower limit logics, a set of premises may entail a disjunction of abnormalities without entailing any of its disjuncts. Consider, for instance, full positive logic plus $A \vee \neg A$ – this is the lower limit logic of the inconsistency-adaptive logic that we shall see below. According to this logic, neither $p \wedge \neg p$ nor $q \wedge \neg q$ is derivable from $\{p, q, \neg p \vee \neg q\}$, but $(p \wedge \neg p) \vee (q \wedge \neg q)$ is.

Below, a disjunction of abnormalities will be called a *Dab*-formula, and will be written as $Dab(\Delta)$, in which Δ is a finite subset of Ω. If $Dab(\Delta)$ is derivable from Γ by the lower limit logic, it will be called a *Dab*-consequence of Γ. $Dab(\Delta)$ will be called a *minimal Dab*-consequence of Γ if and only if it is a *Dab*-consequence of Γ, and there is no $\Delta' \subset \Delta$ such that $Dab(\Delta')$ is a *Dab*-consequence of Γ.

If $Dab(\Delta)$ is a minimal *Dab*-consequence of Γ, then one of the members of Δ behaves abnormally with respect to Γ, but it cannot be established which

one. Hence, the phrase that Γ should be interpreted "as normally as possible" (that is, that it should be assumed for as many abnormalities as possible that they are false) becomes ambiguous. It is in view of this complication that the third element, the adaptive strategy, is needed.

The most common strategies are the Reliability Strategy and the Minimal Abnormality Strategy.[27] According to the former, a formula behaves abnormally with respect to Γ if and only if it is a member of a minimal *Dab*-consequence of Γ. The Minimal Abnormality leads to a somewhat richer consequence set. If, for example, $Dab(\Delta_1)$, ..., $Dab(\Delta_n)$ are the minimal *Dab*-consequences of Γ, then the Minimal Abnormality Strategy first selects one member of each Δ_i to behave abnormally, and next, considers all such combinations.

The difference between the two strategies is best illustrated by means of an example. Consider, for instance, $\Gamma = \{\neg p, \neg q, p \vee q, p \vee r, q \vee r, s, \neg s\}$, and suppose that an inconsistency-adaptive logic is applied to this. If it is based on the Reliability Strategy, r is not an adaptive consequence of Γ (p, q and s are all considered as behaving abnormally). If it is based on the Minimal Abnormality Strategy, then r is an adaptive consequence. The reason is that the Minimal Abnormality Strategy considers the situation in which only p and s behave abnormally, and also the situation in which only q and s behave abnormally, but does not consider the situation in which all three of them behave abnormally. Hence, in all situations considered, r comes out true: if only p and s behave abnormally, then r is true in view of $\neg q$ and $q \vee r$; if only q and s behave abnormally, then r is true in view of $\neg p$ and $p \vee r$.

The semantics of an adaptive logic is obtained by selecting a subset of the models of the lower limit logic. The selection is determined by the adaptive strategy. The Minimal Abnormality Strategy selects those models that are, in a set-theoretical sense, minimally abnormal. The Reliability Strategy selects the models in which no other abnormalities are verified than those that occur as a disjunct of some minimal *Dab*-consequence of the premises. If some theory Γ behaves normally with respect to the upper limit logic, the adaptive models of Γ coincide with the models of the upper limit logic that validate Γ.

Syntactically, an adaptive logic is obtained by considering the inference rules of the lower limit logic as unconditional, and the inference rules of the upper limit logic (that are not validated by the lower limit logic) as conditional. The adaptive strategy determines a marking definition (see below). Sentences derived conditionally at some stage in a proof may at a later stage be rejected, namely, when the condition is no longer satisfied. As we shall see in the next section, the mechanism by which this is realized is quite simple. If a formula is added by the application of a conditional rule, a 'condition' (set of abnormalities) that is specified by the rule, is written to the right of the line. If a formula is added by the application of an unconditional rule, no condition is introduced, but the conditions (if any) that affect the premises of the application are conjoined for its conclusion. The members of the condition have to behave normally for the formula to be derivable. At each stage of the proof – with each formula added – the

marking definition is invoked: for each line that has a condition attached to it, it is checked whether the condition is fulfilled or not. If it is not, the line is *marked*, and the formula that occurs on it is no longer considered as derived.

5. THE GENERIC FORMAT OF THE PROOF THEORY

The proof theory of adaptive logics that satisfy some minimal requirements can be characterized in a generic format.[28] In view of the illustrations in the next section, I briefly discuss this format here.

The motor for the proof theory is provided by the following theorem:

Theorem 1 $B_1,\ldots,B_n \vdash_{\mathbf{ULL}} A$ *if and only if there is a finite* $\Delta \subseteq \Omega$ *such that* $B_1,\ldots,B_n \vdash_{\mathbf{LLL}} A \vee Dab(\Delta)$. *(Derivability Adjustment Theorem)*

Theorem 1 warrants that whenever A is **ULL**-derivable from $B_1, \ldots, B_n$, A is **LLL**-derivable from $B_1, \ldots, B_n$ or certain formulas behave abnormally with respect to $B_1, \ldots, B_n$ (remember that $Dab(\Delta)$ is a disjunction of abnormalities). This naturally suggests that, in the dynamic proofs, we derive B from $A_1, \ldots, A_n$, on the condition that no member of Δ behaves abnormally.

In line with this, the proof theory is characterized by three rules – a premise rule PREM, an unconditional rule RU, and a conditional rule RC – and one marking definition. The proofs themselves look almost like any other proof, except that every line in them has a condition attached to it. Thus, lines in a dynamic proof have five elements: (i) a line number, (ii) the formula A that is derived, (iii) the line numbers of the formulas from which A is derived, (iv) the rule by which A is derived, and (v) the condition. Here are the three rules for a proof from Γ:

PREM If $A \in \Gamma$, then one may add to the proof a line consisting of
- (i) the appropriate line number,
- (ii) A,
- (iii) "–",
- (iv) "PREM", and
- (v) $\varnothing$.

RU If $B_1,\ldots,B_n \vdash_{\mathbf{LLL}} A$ $(n \geq 0)$, and $B_1,\ldots,B_n$ occur in the proof on the conditions $\Delta_1,\ldots,\Delta_n$ respectively, then one may add to the proof a line consisting of:
- (i) the appropriate line number,
- (ii) A,
- (iii) the line numbers of the B_i,
- (iv) "RU", and
- (v) $\Delta_1 \cup \ldots \cup \Delta_n$.

RC If $B_1,\ldots,B_n \vdash_{LLL} A \vee Dab(\Theta)$ $(n \geq 0)$, and $B_1,\ldots,B_n$ occur in the proof on the conditions $\Delta_1,\ldots,\Delta_n$ respectively, then one may add to the proof a line consisting of:
(i) the appropriate line number,
(ii) A,
(iii) the line number of the B_i,
(iv) "RC", and
(v) $\Theta \cup \Delta_1 \cup \ldots \cup \Delta_n$.

As mentioned above, the marking definition depends on the strategy. As the illustrations in the next section require the Reliability Strategy, I shall restrict myself to this strategy.[29]

Intuitively, the Reliability Strategy leads to the marking of a line (at some stage), if one of the elements in its condition is unreliable at that stage – that is, is a disjunct of a minimal *Dab*-formula that has been derived at that stage. Let $U_s(\Gamma)$ be the union of all Δ for which $Dab(\Delta)$ is a minimal *Dab*-formula derived at stage s. Below, the superscript in "$\mathbf{AL}^r$" refers to the Reliability Strategy – I omit it for properties that apply to all adaptive logics regardless of their strategy.

Definition 1 *Marking Definition for* $\mathbf{AL}^r$: *Line i is marked at stage s iff, where* Δ *is its fifth element,* $\Delta \cap U_s(\Gamma) \neq \varnothing$.

In view of the marking definition, two forms of derivability can be defined – derivability at a stage and final derivability:

Definition 2 A is *derived at stage s in an* **AL***-proof from* Γ *iff* A *is the second element of a line that is not marked in the proof (at stage s).*

Definition 3 A is *finally derived on line i of an* **AL***-proof from* Γ *iff (i)* A *is the second element of line i, (ii) line i is not marked in the proof, and (iii) any extension of the proof in which line i is marked may be further extended in such a way that line i is unmarked.*

As may be expected, the derivability relation is defined with respect to final derivability:

Definition 4 $\Gamma \vdash_{AL} A$ *(*A *is finally derivable from* Γ*) iff* A *is finally derived on some line of an* **AL***-proof from* Γ.

The following theorem warrants that the dynamics of the proofs is sensible – 'in the end' different dynamic proofs lead to the same conclusions:

Theorem 2 If $\Gamma \vdash_{AL} A$, *then any proof from* Γ *can be extended into a proof in which A is finally derived from* Γ. *(Proof Invariance.)*

6. SOME ILLUSTRATIONS

In this section, I briefly discuss three different adaptive logics that can be used for the study of specific aspects of scientific reasoning. As the proof theory is most useful with respect to this study, I shall not discuss their semantics. So, in view of the generic format of the proof theory presented in the previous section, and the fact that all three of them are based on the Reliability Strategy, I only have to specify their lower limit logic and their set of abnormalities. I begin with the inconsistency-adaptive logic $\mathbf{P}^r$.[30]

6.1 The Inconsistency-Adaptive Logic $\mathbf{P}^r$

The application context of $\mathbf{P}^r$ concerns theories that were intended to be interpreted in terms of **CL**, but that were discovered to be inconsistent. Applying $\mathbf{P}^r$ to such a theory ensures that the theory is interpreted *as consistently as possible*: one obtains the theory in its full richness, except for the pernicious consequences of its inconsistency. What this comes to is that $\mathbf{P}^r$ leads to a much richer interpretation of the theory than any monotonic paraconsistent logic can provide. The reason is that, in all monotonic paraconsistent logics, triviality is avoided by simply invalidating (in a global way) some rules of inference of **CL**. The logic $\mathbf{P}^r$, however, does not invalidate rules of inference, but only specific *applications* of such rules.

As follows from the above description, $\mathbf{P}^r$ is a corrective system. Hence, its upper limit logic is **CL**. Its lower limit logic is **P** – full positive logic plus $A \vee \neg A$, and its set of abnormalities Ω consists of all formulas of the form $\exists(A \wedge \neg A)$. Note that extending **P** with the presupposition that all members of Ω are false indeed results in the upper limit logic **CL**. Thus, it is warranted that, for every *consistent* Γ, $Cn_{\mathbf{P}^r}(\Gamma) = Cn_{\mathbf{CL}}(\Gamma)$.

As explained in the previous section, the proof theory of adaptive logics is based on a specific relation between derivability by the upper limit logic and derivability by the lower limit logic. This relation is expressed in Theorem 1. The following examples illustrate that theorem in the case of $\mathbf{P}^r$:

(2) $\Gamma_1 = \{\neg t, t \vee s\}$
$\Gamma_1 \vdash_{\mathbf{CL}} s$
$\Gamma_1 \vdash_{\mathbf{P}} s \vee (t \wedge \neg t)$

(3) $\Gamma_2 = \{r \supset s, \neg s\}$
$\Gamma_2 \vdash_{CL} \neg r$
$\Gamma_2 \vdash_{P} \neg r \vee (s \wedge \neg s)$

(4) $\Gamma_3 = \{u \supset p, u \supset \neg p\}$
$\Gamma_3 \vdash_{CL} \neg u$
$\Gamma_3 \vdash_{P} \neg u \vee (p \wedge \neg p)$

(5) $\Gamma_4 = \{p, \neg p \vee q, r \supset \neg q\}$
$\Gamma_4 \vdash_{CL} \neg r$
$\Gamma_4 \vdash_{P} \neg r \vee ((p \wedge \neg p) \vee (q \wedge \neg q))$

Let us now look at a simple propositional example of a $\mathbf{P}^r$-proof. Consider, for instance, $\Gamma = \{p \wedge \neg t, q \wedge (t \vee s), \neg p \vee \neg q, r \supset \neg s, u \supset p, u \supset \neg p\}$. The premise rule enables us to introduce each member of Γ on the empty condition:

1	$p \wedge \neg t$	–	PREM	$\varnothing$
2	$q \wedge (t \vee s)$	–	PREM	$\varnothing$
3	$\neg p \vee \neg q$	–	PREM	$\varnothing$
4	$r \supset \neg s$	–	PREM	$\varnothing$
5	$u \supset p$	–	PREM	$\varnothing$
6	$u \supset \neg p$	–	PREM	$\varnothing$

In view of (4), the rule RC allows us to derive $\neg u$ from the formulas on lines 5 and 6, but on the condition that $p \wedge \neg p$ behaves normally: [31]

...				
7	$\neg u$	5,6	RC	$\{p \wedge \neg p\}$

As $\neg t$ and $t \vee s$ are **P**-derivable from, respectively, the formula on line 1 and that on line 2, both can be added to the proof by means of the rule RU:

...				
8	$\neg t$	1	RU	$\varnothing$
9	$t \vee s$	2	RU	$\varnothing$

In view of (2), RC can be applied to the formulas on lines 8 and 9:

...				
10	s	8,9	RC	$\{t \wedge \neg t\}$

In view of (3), RC can be applied to the formulas on line 4 and 10. As the condition of line 10 is not empty, it is 'carried over', and the condition for the new line is that both $t \wedge \neg t$ and $s \wedge \neg s$ should behave normally:

...				
11	$\neg r$	4,10	RC	$\{t \wedge \neg t, s \wedge \neg s\}$

Up to this stage in the proof, no *Dab*-formula has been derived. However, from lines 1–3, $(p \wedge \neg p) \vee (q \wedge \neg q)$ is **P**-derivable, and hence, may be added by the rule RU:

...				
12	$(p \wedge \neg p) \vee (q \wedge \neg q)$	1–3	RU	$\varnothing$

At this stage of the proof, it is clear that both $p \wedge \neg p$ and $q \wedge \neg q$ are unreliable, and hence, in view of the marking definition, line 7 is marked:

...				
7	$\neg u$	5,6	RC	$\{p \wedge \neg p\}\surd_{12}$
...				
12	$(p \wedge \neg p) \vee (q \wedge \neg q)$	1–3	RU	$\varnothing$

At stage 12, the formula on line 7 is no longer considered as derived. Given the simplicity of the premises, it is easily observed that line 7 will remain marked in any further extension of the proof, and hence, that $\neg u$ is not finally derivable from the premises.[32] The formulas on line 10 and 11 are finally derived. Note especially that the formulas from which the formula on line 10 is derived are themselves 'conjoined' to unreliable formulas (see lines 1 and 2). This illustrates that the adaptation of the logic $\mathbf{P}^r$ does not occur at the level of the individual premises, but at the level of its **P**-consequences.[33]

Some readers may have problems with the fact that line 7 was introduced in the proof while the formula in its condition is obviously unreliable. Note, however, that the example is not a realistic one: with such simple premises hardly any reasoner will need a dynamic proof. But, as soon as the examples become a bit more interesting, even skilled reasoners are unable to see at once all the relevant consequences of the premises. In view of this, and in view of the lack of a positive test for consistency, it is sensible to make the application of the inference rules and of the marking definition relative to the distinctions that, at a certain stage, are made by the reasoner.

6.2 *The Logic of Inductive Generalization* $\mathbf{IL}^r$

The first adaptive logic of inductive generalization, **LI**, was designed by Diderik Batens.[34] This system is intended to be applied to sets of premises that consist, on the one hand, of a number of empirical data (singular statements) and, on the other hand, of a number of background generalizations. From these, it leads to a number of new generalizations (that are *jointly compatible* with the premises) and to the **CL**-consequences of these new generalizations, the data and the background generalizations.

As Diderik Batens and Lieven Haesaert have shown,[35] the logic **LI** can be formulated according to the usual standards for adaptive logics. For pedagogical reasons, their reconstruction proceeds in two steps. First, they present a system $\mathbf{IL}^r$ that can only be applied to sets of empirical data. Next, this system is upgraded to one that can handle both empirical data and background generalizations. The reason for this two-step procedure is that the adequate handling of the background assumptions requires a prioritized adaptive logic. In the same paper, Batens and Haesaert also present some variants for **LI** that overcome several of the simplifying assumptions of the latter.

The aim here is not to present a detailed presentation of any of these systems, but only to give a rough idea of the basic mechanism. Hence, I shall restrict myself to the (flat) adaptive logic $\mathbf{IL}^r$.

The basic idea behind $\mathbf{IL}^r$ is to assume (unless and until proven false) that, if something holds true for some objects, it holds true for all objects. In line with this, the set of abnormalities of $\mathbf{IL}^r$ consists of all formulas of the form $\exists A \wedge \exists \neg A$, in which A is purely functional (that is, no individual constant, sentential letter or quantifier occurs in it). Its lower limit logic is **CL** and its upper limit logic, **UCL**, is obtained by extending **CL** with the axiom $\exists A \supset \forall A$, in which $\forall A$ stands for the universal closure of A. The following examples illustrate Theorem 1 as applied to $\mathbf{IL}^r$ – $Dab_a\{A_1,...,A_n\}$ $(n \geq 1)$ abbreviates $(\exists A_1 \wedge \exists \neg A_1) \vee ... \vee (\exists A_n \wedge \exists \neg A_n)$:[36]

(6) $\Gamma_1 = \{Pa\}$

$\Gamma_1 \vdash_{\mathbf{UCL}} (\forall x)Px$

$\Gamma_1 \vdash_{\mathbf{CL}} (\forall x)Px \vee Dab_a\{Px\}$

(7) $\Gamma_2 = \{Pa, Qa\}$

$\Gamma_2 \vdash_{\mathbf{UCL}} (\forall x)(Px \supset Qx)$

$\Gamma_2 \vdash_{\mathbf{CL}} (\forall x)(Px \supset Qx) \vee Dab_a\{Px \supset Qx\}$

(8) $\Gamma_3 = \{Pa, Qa \vee Ra\}$

$\Gamma_3 \vdash_{\mathbf{UCL}} (\forall x)(Px \supset (Qx \vee Rx))$

$\Gamma_3 \vdash_{\mathbf{CL}} (\forall x)(Px \supset (Qx \vee Rx)) \vee Dab_a\{Px \supset (Qx \vee Rx)\}$

(9) $\Gamma_4 = \{Pa, Qa, Ra\}$
$\Gamma_4 \vdash_{\mathbf{UCL}} (\forall x)(Px \supset (Qx \wedge Rx))$
$\Gamma_4 \vdash_{\mathbf{CL}} (\forall x)(Px \supset (Qx \wedge Rx)) \vee Dab_a\{Px \supset Qx, Px \supset Rx\}$

I now present a simple example of an $\mathbf{IL}^r$-proof. Remember that **CL** is the lower limit logic, and hence, that all **CL**-derivable formulas can be added by means of the rule RU. In the fifth element, A abbreviates $\exists A \wedge \exists \neg A$.

1	$(Pa \wedge Pb) \wedge (Pc \wedge Pd)$	–	PREM	$\varnothing$
2	$(Qa \wedge Qb) \wedge Rc$	–	PREM	$\varnothing$
3	$\neg Qb \vee \neg Rb$	–	PREM	$\varnothing$
4	$Pa \wedge Qa$	1,2	RU	$\varnothing$
5	$Pc \wedge Rc$	1,2	RU	$\varnothing$
6	$Pb \wedge (\neg Qb \vee \neg Rb)$	1,3	RU	$\varnothing$
7	$(\forall x)(Px \supset Qx)$	4	RC	$\{Px \supset Qx\}$
8	$(\forall x)(Px \supset Rx)$	5	RC	$\{Px \supset Rx\}$

The formulas on lines 7 and 8 are justified in view of (7). Suppose that we now continue the proof as follows:

...

9	$(\forall x)(Px \supset (Qx \wedge Rx))$	7,8	RU	$\{Px \supset Qx, Px \supset Rx\}$
10	$Qb \wedge Rb$	1,9	RU	$\{Px \supset Qx, Px \supset Rx\}$

At this point, it may seem that we are in trouble, for the formula on line 10 is clearly incompatible with that on line 3. This, however, only indicates that the generalizations on lines 7 and 8 are not jointly compatible with the premises (given our insight in the premises at this stage of the proof, both generalizations are compatible with the premises, but their conjunction is not). This can also be seen from the fact the following disjunction of abnormalities is **CL**-derivable:

...

11	$Dab_a\{Px \supset Qx, Px \supset Rx\}$	1,2,6	RU	$\varnothing$

The formula on line 11 indicates that either $(\exists x)(Px \supset Qx) \wedge (\exists x)\neg(Px \supset Qx)$ or $(\exists x)(Px \supset Rx) \wedge (\exists x)\neg(Px \supset Rx)$ behaves abnormally. Moreover, as the formula on line 11 is the *shortest* disjunction of abnormalities derived at stage 11, it cannot be decided, at this stage, which one behaves abnormally. Hence, in line with the Reliability Strategy, lines 7–10 are marked:

...

7	$(\forall x)(Px \supset Qx)$	4	RC	$\{Px \supset Qx\}\surd_{11}$
8	$(\forall x)(Px \supset Rx)$	5	RC	$\{Px \supset Rx\}\surd_{11}$
9	$(\forall x)(Px \supset (Qx \wedge Rx))$	7,8	RU	$\{Px \supset Qx, Px \supset Rx\}\surd_{11}$
10	$Qb \wedge Rb$	1,9	RU	$\{Px \supset Qx, Px \supset Rx\}\surd_{11}$
11	$Dab_a\{Px \supset Qx, Px \supset Rx\}$	1,2,6	RU	$\varnothing$

The following generalization and prediction, however, are not affected by the *Dab*-formula on line 11, and hence, can be added without being marked at this stage:

...

12	$(\forall x)(Px \supset (Qx \vee Rx))$	4,5	RC	$\{Px \supset (Qx \vee Rx)\}$
13	$Qd \vee Rd$	1,12	RU	$\{Px \supset (Qx \vee Rx)\}$

Thus far, none of the examples involved the unmarking of previously marked lines. Let us add a few more lines to the proof to illustrate also this side of the dynamics:

...

14	$\neg Rb$	2,3	RU	$\varnothing$
15	$Pb \wedge \neg Rb$	1,14	RU	$\varnothing$
16	$Dab_a\{Px \supset Rx\}$	5,15	RU	$\varnothing$

In view of the formula on line 16, the formula on line 11 is no longer the shortest disjunction of abnormalities. Thus, given the distinctions made at this stage of the proof, it becomes clear that $(\exists x)(Px \supset Rx) \wedge (\exists x)\neg(Px \supset Rx)$ behaves abnormally, and that there is no reason to assume that also $(\exists x)(Px \supset Qx) \wedge (\exists x)\neg(Px \supset Qx)$ behaves abnormally. In line with this, and as is assured by the marking definition, line 7 is *unmarked* at this stage:

...

7	$(\forall x)(Px \supset Qx)$	4	RC	$\{Px \supset Qx\}\surd_{11}$
8	$(\forall x)(Px \supset Rx)$	5	RC	$\{Px \supset Rx\}\surd_{11}$
9	$(\forall x)(Px \supset (Qx \wedge Rx))$	7,8	RU	$\{Px \supset Qx, Px \supset Rx\}\surd_{11}$
10	$Qb \wedge Rb$	1,9	RU	$\{Px \supset Qx, Px \supset Rx\}\surd_{11}$
11	$Dab_a\{Px \supset Qx, Px \supset Rx\}$	1,2,6	RU	$\varnothing$

...

16 $Dab_{a}\{Px \supset Rx\}$ 5,15 RU $\varnothing$

At stage 16 of the proof, the formula on line 7 is again considered as derived. This shows that the dynamics works in both directions: formulas derived at some stage may have to be considered as not derived at a later stage, but may at an even later stage again be considered as derived. All this may sound rather unfamiliar, but it is important to remember that the dynamics is perfectly under control. The dynamic proof theory of $\mathbf{IL}^r$ is sound and complete with respect to a proper semantics.[37] This warrants that the dynamics is bound to lead to stability, and that the stable situation is determined by the premises only.

6.3 The Logic of Abduction $\mathbf{LA}^r$

The first results on abductive adaptive logics were obtained quite recently.[38] These results, however, only concern the dynamic proof theory and are not formulated within the usual adaptive framework. In this section, I briefly show how an adaptive logic for abduction can be obtained that satisfies the usual standards of adaptive logics – the resulting logic will be called $\mathbf{LA}^r$.[39]

The idea behind any adaptive logic for abduction is that it should lead from a set of explananda together with a background theory to a set of explanatory hypotheses as well as to the **CL**-consequences of these and the background theory (the latter serve as predictions). They differ from one another with respect to the constraints that are imposed on the explananda and on the explanatory hypotheses. The latter is related to the fact that abduction may be used for different reasons: explanatory hypotheses may be derived because of a theoretical interest or because of a practical need. In the latter case, but not in the former, one usually requires that the explanatory hypotheses are mutually consistent and that they are 'as weak as possible'. In medical diagnosis, for instance, one typically tries to abduce the weakest possible hypothesis before actions are taken.

As far as the choice of the explananda is concerned, $\mathbf{LA}^r$ only considers singular statements, and hence, only handles the predicative version of abduction, namely

(10) $(\forall \alpha)(A(\alpha) \supset B(\alpha)), B(\beta/\alpha) / A(\beta/\alpha)$

in which $A(\beta/\alpha)$ stands for the formula obtained from $A(\alpha)$ by replacing every free occurrence of the variable α by the constant β.

The intended application context of $\mathbf{LA}^r$ is the situation in which hypotheses are derived as a basis for actions. In line with this, it leads to a set of hypotheses that are jointly compatible with the premises and that are, moreover, as weak as possible in view of them.

For instance, if $\Gamma = \{(\forall x)(Px \supset Qx), (\forall x)(\neg Px \supset Rx), Qa, Ra\}$, neither Pa nor $\neg Pa$ is derivable; if $\Gamma = \{(\forall x)(Px \supset Qx), (\forall x)(Rx \supset Qx), Qa\}$, neither Pa nor Ra is derivable, but $Pa \vee Ra$ is.

Within these constraints, $\mathbf{LA}^r$ validates as many applications of (10) as possible. Thus, it is considered as an abnormality when, for some A, B and β, both $(\forall \alpha)(A(\alpha) \supset B(\alpha))$ and $B(\beta/\alpha)$ are true and $A(\beta/\alpha)$ is false. In line with this, the set of abnormalities of $\mathbf{LA}^r$ consists of all closed formulas of the form $(\forall \alpha)(A(\alpha) \supset B(\alpha)) \wedge \neg(B(\beta/\alpha) \supset A(\beta/\alpha))$. Note that the latter is **CL**-equivalent to $(\forall \alpha)(A(\alpha) \supset B(\alpha)) \wedge (\neg A(\beta/\alpha) \wedge B(\beta/\alpha))$. The upper limit logic of $\mathbf{LA}^r$, say **ACL**, is obtained by extending **CL** with the following axiom:

$$(\forall \alpha)(A(\alpha) \supset B(\alpha)) \supset (\forall \alpha)(B(\alpha) \supset A(\alpha)).$$

As for $\mathbf{IL}^r$, the lower limit logic of $\mathbf{LA}^r$ is **CL**.

As before, I shall represent disjunctions of abnormalities in an abbreviated form. From now on, $Dab_a\{(A_1(\alpha_1) \supset B_1(\alpha_1))^{\beta_1}, \ldots, (A_n(\alpha_n) \supset B_n(\alpha_n))^{\beta_n}\}$ will abbreviate $\vee\{(\forall \alpha_i)(A_i(\alpha_i) \supset B_i(\alpha_i)) \wedge \neg(B_i(\beta_i/\alpha_i) \supset A_i(\beta_i/\alpha_i))\}$ $(1 \leq i \leq n)$.

The following examples illustrate Theorem 1 as applied to $\mathbf{LA}^r$:

(11) $\Gamma_1 = \{(\forall x)(Px \supset Qx), Qa\}$
$\Gamma_1 \vdash_{\mathbf{ACL}} Pa$
$\Gamma_1 \vdash_{\mathbf{CL}} Pa \vee Dab_a\{(Px \supset Qx)^a\}$

(12) $\Gamma_2 = \{(\forall x)(Px \supset Qx), (\forall x)(Qx \supset Rx), Ra\}$
$\Gamma_2 \vdash_{\mathbf{ACL}} Pa$
$\Gamma_2 \vdash_{\mathbf{CL}} Pa \vee Dab_a\{(Px \supset Qx)^a, (Qx \supset Rx)^a\}$

(13) $\Gamma_3 = \{(\forall x)(Px \supset Qx), (\forall x)(Rx \supset Sx), Qa, Sa\}$
$\Gamma_3 \vdash_{\mathbf{ACL}} Pa \wedge Ra$
$\Gamma_3 \vdash_{\mathbf{CL}} (Pa \wedge Ra) \vee Dab_a\{(Px \supset Qx)^a, (Rx \supset Sx)^a\}$

(14) $\Gamma_4 = \{(\forall x)(Px \supset Qx), (\forall x)(Rx \supset Qx), Qa\}$
$\Gamma_4 \vdash_{\mathbf{ACL}} Pa \vee Ra$
$\Gamma_4 \vdash_{\mathbf{CL}} (Pa \vee Ra) \vee Dab_a\{(Px \supset Qx)^a, (Rx \supset Qx)^a\}$

Let me now illustrate the main properties of $\mathbf{LA}^r$ with a simple example of an $\mathbf{LA}^r$-proof. Where $(A(\alpha) \supset B(\alpha))^\beta$ occurs in the fifth element, it abbreviates $(\forall \alpha)(A(\alpha) \supset B(\alpha)) \wedge \neg(B(\beta/\alpha) \supset A(\beta/\alpha))$.

1	$(\forall x)(Px \supset Qx)$	–	PREM	$\varnothing$
2	$(\forall x)(Rx \supset Qx)$	–	PREM	$\varnothing$
3	$(\forall x)(Sx \supset \neg Px)$	–	PREM	$\varnothing$

4	$Qa \wedge Sa$	–	PREM	$\varnothing$
5	$Qb \wedge (\neg Pb \vee \neg Rb)$	–	PREM	$\varnothing$
6	Qc	–	PREM	$\varnothing$
7	Qa	4	RU	$\varnothing$
8	Pa	1,7	RC	$\{(Px \supset Qx)^a\}\surd_{10}$
9	$\neg Pa$	3,4	RU	$\varnothing$
10	$Dab_a\{(Px \supset Qx)^a\}$	1,7,9	RU	$\varnothing$

At line 8, and in view of (11), Pa is derived as an explanatory hypothesis for Qa. As appears from line 9, this hypothesis is not compatible with the premises – its negation is **CL**-derivable from them. This, however, implies that $(\forall x)(Px \supset Qx) \wedge \neg(Qa \supset Pa)$ is **CL**-derivable from the premises. As soon as this formula is added to the proof (see line 10), line 8 is marked and will remain marked in any extension of the proof. Hence, the hypothesis Pa is not finally derivable from the premises. The following continuation of the proof illustrates what happens if two hypotheses are not mutually compatible with the premises:

11	Qb	5	RU	$\varnothing$
12	Pb	1,11	RC	$\{(Px \supset Qx)^b\}\surd_{16}$
13	Rb	2,11	RC	$\{(Rx \supset Qx)^b\}\surd_{16}$
14	$Pb \wedge Rb$	12,13	RC	$\{(Px \supset Qx)^b, (Rx \supset Qx)^b\}\surd_{16}$
15	$\neg Sb$	3,12	RU	$\{(Px \supset Qx)^b\}\surd_{16}$
16	$Dab_a\{(Px \supset Qx)^b, (Rx \supset Qx)^b\}$	1,2,5	RU	$\varnothing$

As for $\mathbf{IL}^r$, $\mathbf{LA}^r$ enables one to derive different hypotheses and to conjoin these – see lines 12–14. It also enables one to derive predictions from the hypotheses and the premises (see line 15). However, if a series of hypotheses is not mutually consistent with the premises, the Reliability Strategy warrants that all of them (together with their consequences) are marked. In this case, it is **CL**-derivable from the premises that $(\forall x)(Px \supset Qx) \wedge \neg(Qb \supset Pb)$ or $(\forall x)(Rx \supset Qx) \wedge \neg(Qb \supset Rb)$ behaves abnormally, and hence, lines 12–15 are marked as soon as this formula is added to the proof (see line 16).

Let us now see what happens if more than one explanatory hypothesis can be abduced for the same explanandum:

17	Pc	1,6	RC	$\{(Px \supset Qx)^c\}\surd_{22}$
18	Rc	2,6	RC	$\{(Rx \supset Qx)^c\}\surd_{21}$
19	$(\forall x)((Px \vee Rx) \supset Qx)$	1,2	RU	$\varnothing$
20	$Pc \vee Rc$	6,19	RC	$\{((Px \vee Rx) \supset Qx)^c\}$

21	$Dab_a\{(Px \wedge \neg Rx) \supset Qx)^c, (Rx \supset Qx)^c\}$	1,2,6	RU	$\varnothing$
22	$Dab_a\{(Rx \wedge \neg Px) \supset Qx)^c, (Px \supset Qx)^c\}$	1,2,6	RU	$\varnothing$

At lines 17, 18 and 20, three different explanatory hypotheses are derived for Qc. However, in view of the *Dab*-formulas on lines 21 and 22, only the logically weakest one is retained.

Some readers may wonder whether this mechanism is also put into use if the antecedent of some 'covering law' is arbitrarily strengthened – a problem that is well-known from the literature on abduction.[40] Let me first illustrate the problem:

23	$(\forall x)((Px \wedge Tx) \supset Qx)$	1	RU	$\varnothing$
24	$(\forall x)((Px \wedge \neg Tx) \supset Qx)$	1	RU	$\varnothing$
25	$Pc \wedge Tc$	7,23	RC	$\{((Px \wedge Tx) \supset Qx)^c\}$
26	$Pc \wedge \neg Tc$	7,24	RC	$\{((Px \wedge \neg Tx) \supset Qx)^c\}$

If the formulas on lines 25 and 26 would be finally derivable from the premises, one would not only obtain explanatory hypotheses that are partly irrelevant, but even triviality:

27	Tc	25	RU	$\{((Px \wedge Tx) \supset Qx)^c\}$
28	$\neg Tc$	26	RU	$\{((Px \wedge \neg Tx) \supset Qx)^c\}$
29	p	27,28	RU	$\{((Px \wedge Tx) \supset Qx)^c, ((Px \wedge \neg Tx) \supset Qx)^c\}$

However, as the formulas on lines 25 and 26 are not jointly compatible with the premises, the following *Dab*-formula is **CL**-derivable:

30	$Dab_a\{((Px \wedge Tx) \supset Qx)^c, ((Px \wedge \neg Tx) \supset Qx)^c\}$	1,6	RU	$\varnothing$

As soon as this line is added to the proof, the marking definition warrants that lines 25–29 are marked. It is, moreover, easily observed that these lines will remain marked in any further extension of the proof, and hence, that the formulas on these lines are not finally derivable.

7. IN CONCLUSION

In this paper, I argued that it is typical of many scientific reasoning processes that they involve an integrated use of induction and deduction, and showed how and why this gives rise to dynamic forms of reasoning. I also argued that the resulting reasoning processes can adequately be explicated in terms of adaptive logics.

There seem to be several reasons why this explication is important. First, it enables one to study a variety of reasoning processes (from reasoning from inconsistencies over abductive reasoning to analogical reasoning) in a unified framework. Next, it enables one to reconstruct, in a formally exact and yet realistic way, actual cases of scientific reasoning. Finally, it provides a better insight in the *appraisal* of 'dynamic arguments'. For instance, it not only teaches that the conclusion of such an argument can sometimes be undermined by simply continuing it, but also delineates which kind of consequences will have this effect.

Still, there may be several kinds of objections. A first one may be that the study of dynamic reasoning patterns is not a matter of logic, but of heuristics. This is, for instance, the line followed by Hintikka and by those who applied his interrogative model.[41] As these applications show, it is indeed possible, given the proper kinds of heuristic moves, to capture various forms of dynamic reasoning (such as the generation of hypotheses and of questions that are sensible in an explanation-seeking process) within the framework of **CL**. An important argument in favour of a logic-based approach, however, is that it provides an insight in the *soundness* of dynamic reasoning processes. If, for instance, in situations for which the logic $\mathbf{LA}^r$ is adequate, an explanatory hypothesis is *not* withdrawn when it has been established to be incompatible with the data, then this is not classified as a case of bad heuristics, but as a case of *bad reasoning*.[42]

A second objection may be that adaptive logics simply reconstruct whatever type of reasoning one finds in the sciences or in everyday situations, and hence, have no normative impact. Several remarks are in order here. It indeed holds true that the design of an adaptive logic starts from considering examples of actual reasoning processes and/or from one's intuitions about the application context at issue. This, however, is not different for any other logic. Even Classical Logic is not god-given, but is a human artefact that incorporates several prejudices of the time in which it was designed.[43] What is important, however, is that for adaptive logics, just as for any other logic, these examples and intuitions are only a *starting point*. In order to design a decent logic, that has a proper metatheory, the appropriate tools have to applied. And here, frustrating as it sometimes may be, it is not the case that *anything goes*. This is why adaptive logics always led, and most probably always will lead, to a revision of the intuitions they started from.

A last objection may concern the role of empirical constraints in the design of adaptive logics. Some readers may have problems with the fact that the logics are not 'user-neutral': they are designed for a specific type of users (that have, for instance, limited brains and a limited cognitive capacity). This, however, seems to be one of the great challenges formal logicians face when turning to the study of induction. Either they design systems that can only be applied by gods and angels or they take into account some of the characteristics of actual reasoners. At least to me, the problems concerning scientific reasoning seem too important and too interesting not to take the second road.

NOTES

† The author is a Postdoctoral Researcher of the Fund for Scientific Research – Flanders (Belgium) and is indebted to Matti Sintonen for interesting questions and comments on a previous version.

* The research for this paper was supported by the Fund for Scientific Research – Flanders (Belgium), and indirectly by the Flemish Minister responsible for Science and Technology (contract BIL01/80). Unpublished papers referred to in this paper are available at http://logica.UGent.be/centrum/writings.

1. The premises of an individual inference in a reasoning process need not form part of the background premises. Often, the former are themselves conclusions arrived at by previous inferences.
2. We shall see such an example from the history of the sciences in Section 2.
3. See, for instance, Thomas Nickles, "What is a Problem that we may Solve It?", in: *Synthese* 47, 1981, pp. 85-118, Paul Thagard, *Computational Philosophy of Science*. Cambridge, MA: MIT Press/Bradford Books 1988, Lindley Darden, *Theory Change in Science: Strategies from Mendelian Genetics*. Oxford: Oxford University Press 1991, Ronald Giere, *Explaining Science: A Cognitive Approach*. Chicago: University of Chicago Press 1988, Nancy Nersessian, *Faraday to Einstein: Constructing Meaning in Scientific Theories*. Dordrecht: Martinus Nijhoff Publishers 1984, and David Gooding, *Experiment and the Making of Meaning*. Dordrecht: Kluwer 1990. Jaakko Hintikka and his associates, especially Matti Sintonen, were among the few philosophers of science that kept trying to apply logic – see also below.
4. A reasoning process that exhibits an internal dynamics is not necessarily non-monotonic. It has been shown, for instance, that the pure logic of relevant implication can be characterized by a dynamic proof theory (see Diderik Batens, "A Dynamic Characterization of the Pure Logic of Relevant Implication", in: *Journal of Philosophical Logic* 30, 2001, pp. 267-280).
5. Among other things, I mean by this that it should be possible to formulate the metatheory in precise terms.
6. For an analysis of this particular reasoning process, see Joke Meheus, Liza Verhoeven, Maarten Van Dyck & Dagmar Provijn, "Ampliative Adaptive Logics and the Foundation of Logic-Based Approaches to Abduction", in: Lorenzo Magnani, Nancy Nersessian & Claudio Pizzi (eds.), *Logical and Computational Aspects of Model-Based Reasoning*. Dordrecht: Kluwer 2002, pp. 39-71.
7. See Thomas Kuhn, *The Essential Tension*. Chicago: University of Chicago Press 1977, pp. 167-171.
8. Sadi Carnot, *Réflexions sur la Puissance Motrice du Feu et sur les Machines propres à développer cette Puissance*. Paris: Bachelier 1824.
9. For a global reconstruction of this particular episode in the history of the sciences, see Peter Clark, "Atomism versus Thermodynamics", in: Colin Howson (ed.), *Method and Appraisal in*

the Physical Sciences. The Critical Background to Modern Science, 1880–1905. Cambridge: Cambridge University Press 1976, pp. 41-105, and Joke Meheus, "Clausius' Discovery of the First Two Laws of Thermodynamics. A Paradigm of Reasoning from Inconsistencies", in: *Philosophica* 63, 1999, pp. 89-117 (appeared in 2001). An important difference between these two reconstructions concerns the question whether Clausius actually reasoned from an inconsistent set of premises. In the former, the position seems to be that Clausius first resolved the inconsistencies, on the basis of non-logical grounds, and only then continued reasoning. In the latter, it is argued that localizing and resolving the inconsistencies required reasoning from an inconsistent set of premises.

10. Rudolf Clausius, "Ueber die bewegende Kraft der Wärme und die Gesetze, welche sich daraus für die Wärmelehre selbst ableiten lassen", reprinted in: Max Planck (ed.), *Ueber die bewegende Kraft der Wärme und die Gesetze, welche sich daraus für die Wärmelehre selbst ableiten lassen*. Leipzig: Verlag von Wilhelm Engelmann 1989, pp. 1-52.
11. See Rudolf Clausius, "Ueber die bewegende Kraft der Wärme und die Gesetze, welche sich daraus für die Wärmelehre selbst ableiten lassen", revised edition, reprinted and translated in: F. Folie (ed.), *Théorie Mécanique de la Chaleur par R. Clausius*. Paris: Librairie scientifique, industrielle et agricole 1868, pp. 17-106.
12. See Joke Meheus, "Adaptive Logic in Scientific Discovery: the case of Clausius" in: *Logique et Analyse* 165-166, 1993, pp. 359-389 (appeared in 1996) and Joke Meheus, "Inconsistencies in Scientific Discovery. Clausius's Remarkable Derivation of Carnot's Theorem", in: Helge Krach, Geert Vanpaemel & Paul Marage (eds.), *History of Modern Physics.*, Turnhout: Brepols 2002, pp. 143-154.
13. For the purposes of the present paper, it is not important to see what D and E stand for. The concrete form of the proofs can be found in Meheus, "Inconsistencies in Scientific Discovery. Clausius's Remarkable Derivation of Carnot's Theorem", *op. cit.*
14. Some readers may have problems with this. Can one sensibly reason from an inconsistent set of premises, and thus, commit oneself to the simultaneous acceptance of inconsistent statements? It is important to remember, however, that the truth of a statement A not only depends on the state of the world, but also on the language L that is used to express A, and on the correspondence relation R that links L to the world. Evidently, no matter how the world looks like, it is possible to design a language and a correspondence relation such that the true description of the world is inconsistent (see also Diderik Batens, "In Defence of a Programme for Handling Inconsistencies", in: Joke Meheus (ed.), *Inconsistency in Science*. Dordrecht: Kluwer 2002, pp. 129-150). Moreover, as the history of the sciences teaches, this is not merely a theoretical possibility.
15. A logic is called paraconsistent if it does not validate *Ex Falso Quodlibet* – to derive B from A and $\neg A$.
16. In line with this, almost all (monotonic) paraconsistent logics invalidate *Reductio ad Absurdum*. The only exception known to me is the logic **AN** (see Joke Meheus "An Extremely Rich Paraconsistent Logic and the Adaptive Logic Based on It", in: Diderik Batens, Chris Mortensen, Graham Priest & Jean Paul Van Bendegem (eds.), *Frontiers of Paraconsistent Logic. Baldock: Research Studies Press* 2000, pp. 189-201).
17. In Section 6, we shall see that the suggested replacement is not sophisticated enough to lead to an adequate logic. This simplified version is instructive, however, for the argument in the present section.
18. The consequence relation of **CL** is undecidable, but it has a positive test. Hence, if $\Gamma \vdash_{CL} A$, there is bound to 'exist' a **CL**-proof of A from Γ, even if we may never find it.
19. Diderik Batens "Dynamic Dialectial Logics" in: Graham Priest, Richard Routley & Jean Norman (eds.), *Paraconsistent Logic. Essays on the Inconsistent*. München: Philosophia Verlag 1989, pp. 187-217.
20. Consequence relations that proceed in terms of maximal consistent subsets were first proposed by Rescher and Manor (Nicholas Rescher & Ruth Manor, "On Inference from Inconsistent Premises" in: *Theory and Decision* 1, 1970, pp. 179-217), and are today very popular in AI-applications. Reconstructing these consequence relations in terms of adaptive logics has the advantage that it thus becomes possible to design a proof theory for them.

21. See Joke Meheus, "Deductive and Ampliative Adaptive Logics as Tools in the Study of Creativity", in: *Foundations of Science* 4, 3, 1999, pp. 325-336.
22. See http://logica.UGent.be/adlog/albib.html for an overview of the results on adaptive logics.
23. What is meant with the ambiguous term 'part' will become clear below.
24. See also Meheus, "Adaptive Logic in Scientific Discovery: the Case of Clausius", *op. cit.*
25. Prioritized adaptive logics have important applications in the analysis of discussions and their dynamics – see, for instance, Diderik Batens, "Aspects of the Dynamics of Discussions and Logics Handling Them." forthcoming. The formal machinery of prioritized adaptive logics is a bit more complicated than that of flat ones, but not fundamentally different.
26. I follow the characterization of adaptive logics as presented in Diderik Batens, "The Need for Adaptive Logics in Epistemology", forthcoming.
27. Occasionally, the two strategies coincide and may be replaced by a simpler strategy. This is the case if, for every minimal *Dab*-consequence $Dab(\Delta)$ of every premise set Γ, Δ is a singleton.
28. One of the requirements is that the language of the lower limit logic should be rich enough to express that some formula behaves normally, respectively abnormally. I refer to Batens, "The Need for Adaptive Logics in Epistemology", *op. cit.* for further details and for the proofs to the theorems listed below.
29. The rules PREM, RU and RC are the same if another strategy is applied. The marking definition for the Minimal Abnormality Strategy is a bit more complicated than that for the Reliability Strategy.
30. For a presentation of the logic $\mathbf{P}^r$ at the predicative level, see Diderik Batens, "Inconsistency-adaptive Logics", in: Ewa Orłowska, *Logic at Work. Essays Dedicated to the Memory of Helena Rasiowa*. Heidelberg: Physica Verlag 1999, pp. 445-472. In many papers on the subject, the logic is called "**ACLuN1**".
31. To illustrate the dynamical character of the proofs, I shall each time give the complete proof, but, for reasons of space, omit lines that are not central to see the dynamics.
32. If $q \wedge \neg q$ would be unconditionally derived at some stage in the proof, then $p \wedge \neg p$ would no longer be unreliable (it would not be a disjunct of a *minimal Dab*-formula at that stage), and hence, line 7 would be unmarked. This, however, is the only way to unmark line 7.
33. It follows from this that the logic $\mathbf{P}^r$ is not sensitive to the formulation of the premises, as is the case, for instance, for consequence relations that proceed in terms of maximal consistent subsets. Compare also with footnote 23, to see what was meant there by the ambiguous 'parts'.
34. Diderik Batens, "On a Logic of Induction", forthcoming.
35. Diderik Batens & Lieven Haesaert, "On Classical Adaptive Logics of Induction", to appear.
36. I use this abbreviation instead of the usual $Dab(\Delta)$ to keep the examples within the margins.
37. See Batens & Haesaert, *op. cit.*
38. See Meheus, Verhoeven, Van Dyck & Provijn, *op. cit.* and Joke Meheus " Empirical progress and ampliative adaptive logics", to appear.
39. The results presented in this section concern work in progress. At this moment, the soundness and completeness proofs have not been formulated. However, in view of the format of the proof theory presented here, there is reason to believe that these proofs will be analogous to that for other adaptive logics.
40. It is in view of this problem that **CL**-based approaches to abduction have to require that the explanatory hypothesis is as parsimonious as possible – for a discussion of this problem and its possible solutions, see Atocha Aliseda, *Seeking Explanations. Abduction in Logic, Philosophy of Science and Artificial Intelligence*. Amsterdam: Institute for Logic, Language and Computation 1997.
41. See, for example, Jaakko Hintikka, "What is Abduction? The Fundamental Problem of Contemporary Epistemology." In: *Transactions of the Charles S. Peirce Society* 34, 1998, pp. 503-533, Jaakko Hintikka, Ilpo Halonen, & Arto Mutanen, "Interrogitive Logic as a General Theory of Reasoning", in: Jaakko Hintikka (ed.), *Inquiry as Inquiry: A Logic of Scientific Discovery*. Dordrecht: Kluwer 1999, pp. 47-90 and Matti Sintonen, "In Search of Explanations: From Why-Questions to Shakespearean Questions", in: *Philosophica* 51, 1993, pp. 55-81.
42. The difference between the two approaches is discussed at some length in Diderik Batens & Joke Meheus, "On the Logic and Pragmatics of the Process of Explanation", in: Mika Kiikeri &

Petri Ylikoski (eds.), *Explanatory Connections. Electronic Essays Dedicated to Matti Sintonen.* http://www.valt.helsinki.fi/kfil/matti/, 2001, 22 p.

43. See also Batens, "The Need for Adaptive Logics in Epistemology", *op. cit.*

Centre for Logic and Philosophy of Science
Ghent University
Blandijnberg 2
9000 Gent
Belgium
Joke.Meheus@UGent.be

MATTI SINTONEN

ARGUMENT, INFERENCE AND REASONING – INTEGRATING INDUCTION AND DEDUCTION

Comments on Joke Meheus's "Adaptive Logics and the Integration of Induction and Deduction"

1. LOGIC AND SCIENTIFIC LIFE

In the middle of a conference on the logic of science, an eminent biologist once said: "Does it not bother you guys that we scientists do not use any logic at all." This statement was meant to be a friendly provocation, but there also was a serious message. Scientists often say that the logical analyses are exercises in formal logic and fail to illuminate what the scientists are doing, actual scientific practice. This recurring complaint, although not completely as I will suggest, has not gone unnoticed in the philosophy of science. Indeed, the current trend in analytic philosophy of science as well as in teaching the method of science (if there is such an animal) has been away from (formal) logic as a means of illuminating scientific inquiry.[1]

If this is the spirit in much recent philosophy of science, Joke Meheus and her colleagues in Ghent are refreshing dissidents. They are swimming against the current by arguing that logic still has much to offer for the understanding of scientific reasoning. Meheus argues that what has gone wrong in attempts to appreciate real thinking and real inferences is not that reason has gone on holiday, nor that logic has not been used. Rather, she claims scientific life is amenable to rational reconstruction and, indeed, by logical tools.[2] What needs to be appreciated, she argues, is that the bounds of classical logic are not the bounds of logic or rationality: we need logic, but the logic must be adequate to the task. This need can be appreciated in the case of inductive generalization and abductive-explanatory inference, but most clearly in the case of inconsistencies. Meheus's strategy is to admit that scientists sometimes face situations which force them to entertain inconsistent beliefs, and thus situations which force students of the logic of science to examine the limits of logic and rationality.

Adaptation is a key to survival and success, and it is a laudable enterprise to try and see what adaptive logic could do towards illuminating the cognitive and social enterprise of knowledge. In this paper I shall first try to make it clear what sort of a claim Meheus is putting forth. I shall start by looking at some of the examples (section 2) of adaptive logic in work at the abductive pattern for rea-

F. Stadler (ed.), Induction and Deduction in the Sciences, 121–133.
© 2004 *Kluwer Academic Publishers.*

soning in the case of a singular anomalous phenomenon as well as at the case of Clausius, and the development of electrodynamics. I shall then raise the question as to how urgent the need for para-consistent or adaptive logic is, given that construing scientists as accepting inconsistencies involves some interpretative difficulties (section 3). This brings me to the problem of inconsistencies and Clausius's derivation of Carnot's theorem (section 4) and the problem of scientific discovery based on inconsistent constraints. I shall then distinguish between logical arguments on one hand and inferences and reasoning, on the other, which are geared to using logical arguments in belief formation (5). Section (6) focuses on one motivation behind adopting adaptive logic, viz. faithfulness to scientists' actual reasoning. But how do scientists reason and how do we know how they reason? I shall conclude (7) on the form of the argument here. Meheus is advocating, adaptive, logic in the name of (philosophical) consiliense: the logic is well-behaved and captures something left out by classical logic. The motivation of the analysis seems to be that adaptive logic is one candidate for logic of science in the traditional sense of logic.

2. FROM INCOMPLETENESS TO INCONSISTENCY

It is an axiom in philosophy of science that empirical inquiry proceeds under the assumption that knowledge is incomplete. It is equally a commonplace that scientists confront singular or general facts which are contrary to their theory-based expectations. The two types of case are of course extremely important for the dynamics of science: filling in gaps in the available theories is a moving force in science, and so is that of eliminating anomalies which threaten theories with inconsistencies. Needless perhaps to say, perceiving gaps presupposes a certain view of background knowledge in which theories have a well-defined domain of responsibility so that gaps can be identified as unfinished business. Outright contradictions provide a clearer case: we are taught to shun them; they are a threat to cognitive integrity.

What sort of thing is adaptive logic and what does the claim of integrating deduction and induction amount to? Is the claim that deduction and induction simply cannot be priced apart, and that what counts as deductive or inductive inference depends on the context, in the precise sense detailed by adaptive logic? Or is the claim that although deduction and induction are distinct patterns of *argument* they are used, in actual everyday and especially scientific *inference* and *reasoning*, in tandem? The former claim would be about the same order of radicalness as the claim that the borderline between the analytic and the synthetic is, if it exists at all, drawn by the reason for its own convenience.

Here one is faced with a surprising fact (or with a number of surprising facts). Meheus offers as an example the way these two types of steps combine in the seeking of an explanation: "a novel fact, for instance, explanatory hypotheses are often formulated on the basis of abductive steps. These hypotheses, however,

are withdrawn when their negation is (deductively) derived from the background theory." Her example is the explanation-seeking process, indeed, the abductive process of reasoning, which led Herschel to the discovery of Uranus. Having noted that there was an unexpectedly large object in the vicinity of Zeta Tauri, he first came up with the hypothesis that the large body was a comet. However, this hypothesis did not conform to the observational data about its orbit. As Meheus puts it, the hypothesis was therefore withdrawn and a new one formulated (to the effect that the body was a planet). This insight, just as the first hypothesis, was an outcome of an abductive inferential step.

The crucial question now is: could this chain of reasoning be given reconstruction in terms of classical logic, or rather, inferences sanctioned by classical logic, or do we need to go adaptive? Another question, to be addressed once the more troublesome case of Clausius has been examined, is which one of these logics is the more plausible one. To set the stage in the informal way, Herschel was working within an established Newtonian pattern which enabled him to identify the object as a candidate for comethood. Part of the reason why the comet-hypothesis was embraced was that it indeed fitted the non-theoretical characterization of being a comet. However, this classification did not stand up to subsequent observations and the phenomenon was subsumed under the neighbouring class of intended applications of Newtonian mechanics, the sun-comet-systems.

Now there indeed seem to be alternating inductive and deductive steps here. But how far-reaching conclusions are we to draw from this? We encounter phenomena which either contradict our beliefs (or, more often, phenomena which are unlikely, given our expectations). For classical logic, a small contradiction is as bad as a big one. Should we therefore resort to adaptive logic every time there is trouble, or only when the discrepancy is serious enough to threaten some fundamental commitments? Is the suggestion by Meheus, that adaptive logic is needed to see that discovery can be, rational?

The view that discovery of new hypotheses is not blind or completely blind is not a novel one. In fact, the historicist critics of the received view (even before the "friends of discovery") made it eminently clear that scientists do reason in the context discovery. One of the best examples on the constructive side was Imre Lakatos whose idea of scientific research was based on research programmes. There are some core assumptions which help to point to where the heart of the follower of the research programme should be. They provide the crucial identity of the theory. Apart from the core, there is the protective belt of further assumptions which no doubt can also be ordered vis-à-vis their importance, as well as the positive heuristics. In Lakatos's well-known view, a positive heuristics guides the search of more specific hypotheses which could meet with the internal or external challenges provided by deriving new consequences from the theory and by discovering new observations or experimental results which contradict the theory.

Lakatos (1970, 100) describes this line in an imaginary, but for our discussion, perfect example of "planetary misbehaviour": here a pre-Einsteinian physicist employs Newton's mechanics and its first special law, the law of gravitation *N*, to a newly observed planet. However, given the initial conditions *I* the observations are not what the theory would let him expect. Lakatos writes:

> Does our Newtonian physicist consider that the deviation was forbidden by Newton's theory and therefore that, once established, it refutes the theory *N*? No. He suggests that there must be a hitherto unknown planet *p'*, which perturbs the path of *p*. He calculates the mass, orbit, etc. of this hypothetical planet and then asks an experimental astronomer to test his hypothesis. The planet *p'* is so small that even the biggest available telescopes cannot possibly observe it; the experimental astronomer applies for a research grant to build yet a bigger one. In three years time, the new telescope is ready. Were the unknown planet *p'* to be discovered, it would be hailed as a new victory of Newtonian science. But it is not. Does our scientist abandon Newton's theory and his idea of the perturbing planet? No. He suggests that a cloud of cosmic dust hides the planet from us...[3]

Eventually the hypothesis may be given up, but not because it is simply inconsistent with *N*. This is the sort of dynamics of the career of an empirical theory which already William Whewell recognized: theories do not die because they are inconsistent with facts but because the initial simplicity of the theory crumbles under the new developments.[4]

3. THEORIES, BELIEFS AND COMMITMENTS

The example shows that when problems, empirical or conceptual, arise the scientist is not helpless and therefore need not panic: scientists have reasoned, even informed, guesses as to what the problem is. Needless to say, the alternatives cannot be narrowed down to one, nor can there be guarantee that harmony or consistency could be restored without drifting into the acceptance of increasingly baroque and possibly *ad hoc* maneouvres.

On a Lakatosian rational reconstruction, Herschel was not reasoning from inconsistent knowledge into his new hypotheses. Rather, upon receiving the new observational evidence, he must have realized that something was wrong. Once he did this he had to give up (or rather modify) something in his background knowledge, e.g., by giving up the belief that the object was a comet and then look for the best explanation for it – which, given the evidence, was that it was a planet.[5]

Much depends on what counts as a theory, and what it is to accept a theory or even a theoretical hypothesis. Meheus writes that "hypotheses are withdrawn when their negation is (deductively) derived from the background theory". But the observations which refuted the comet hypothesis did not refute the Newtonian theory and the law of gravitation, or even its particular application to the sun-comet-system. The Lakatosian story (and this is equally true of Kuhn's account, and of many other metatheoritic construals) is based on the assumption

that scientific theories are historically evolving and often hierarchical structured entities. As the structuralist philosophers of science especially have been keen to point out, holding or adopting a theory amounts to using the theory to deal with so-far unexamined applications. It is then natural to distinguish between believing that a theory is true and accepting a theory as worth developing, or entertaining a theory as plausible. It is of course possible that a scientist who adopts this attitude towards a research programme formed around some fundamental laws and further assumptions also believes that some variant of the theory, a theoretically enhanced and more specific proposal, is a true one. But as Bas van Fraassen and Larry Laudan, among others, have proposed, a scientist's primary commitment is to the fundamental theory or the core, not to any specific variant.[6]

4. THE DERIVATION OF CARNOT'S THEOREM

So far the argument has been that construing scientists as believing or accepting inconsistencies involves both historical stage setting and contestable assumptions about what a theory is and what it is to accept or believe in a theory. However, the second example, Clausius' derivation of Carnot's theorem, is even more challenging and much more alarming for someone to whom classical logic is dear. One reason is that at least in physics, but arguably also in the life sciences, it seems uncontestable that a fact counts as a particularly serious anomaly if it is a general phenomen which can be produced experimentally and at will. Of course, a heavenly body which defies our descriptive and explanatory apparatus is an anomaly. But an experimentally confirmed generalization which contradicts a theory is a more serious threat still. Both can be taken to involve abduction since both are ampliative or non-truth-preserving, and indeed, in Meheus's view, both involve reasoning patterns which are conditional. Another reason for anxiety is that Clausius's reasoning which resulted in the new theory relied on two incompatible approaches, those of Carnot and Joule. What is more, if Meheus is right, one could not argue that Carnot simply dropped some problematic parts of one theory to form a consistent set of beliefs, and then expanded the theory. Rather, Clausius had to use both of the incompatible theoretical approaches to derive Carnot's theorem. On top of all, Meheus is not just speaking about two *approaches* being incompatible with one another, but she manages to pick out particular ingredients of theories, and specific beliefs about these ingredients, which contradict each other.

According to Carnot's widely accepted theory, work is a result of transfer of heat, a caloric substance, from a hotter container to a colder one. Although a difference in temperature can be used to do work, sheer transfer in immediate contact does not. To further illustrate the relationship between temperature and work, Carnot described an engine, the Carnot-engine, which can produce work (by using the flow from the hot to the cold), but which can also consume work (when the flow is from the cold to hot). On the assumption that the amount of

caloric given by one reservoir is identical to the one received by the other, Carnot derived the theorem:

(C) It is impossible to design an engine that produces more work than a Carnot-engine, while operating at the same temperatures and absorbing the same amount of heat (from a hot reservoir).[7]

However, this theory flies on the face of Joule's approach in which work in a heat engine results, not from transfer, but from the *conversion* of heat into work. Moreover, this conversion was thought by Joule to be reversible: work and heat are feely convertible into each other.

Against the background of this clash, and the several inconsistencies which Meheus details, Clausius designed two different proofs for Carnot's theorem, starting from the negation of the assumption that Carnot's theorem is true. Now Meheus's analysis draws on the seeming formal parallel between the two arguments using the *Reductio ad Absurdum* pattern. The surprising fact nevertheless is that Clausius thought that only one of them carried force and warranted the inference to Carnot's theorem. The reason, as Meheus construes it, was that in the first derivation the conclusion was inconsistent with the background premises. Here the hypothesis (that Carnot's theorem is false) was idling, and the proof, therefore, failed to address its acceptability. The second proof, in contradistinction, uses the hypothesis in the derivation of the contradiction and, therefore, is, intuitively speaking, relevant for assessing the truth or acceptability of the hypothesis.

5. SCIENTIFIC DISCOVERY: LUCK OR REASONING?

I have thought it necessary to dwell on the case of Clausius and to spell out the theorem somewhat to highlight the news. On the face of it, empirical theories are bound to bump into unexpected observations, and so far undetected and unwanted consequences are a constant possibility. It is far from obvious that even in cases where scientists have no difficulty detecting a contradiction between approaches or theories, they hold that they need to infer into a contradictory set of beliefs. Modern physics, especially the rise of quantum theories, provide good examples. Werner Heisenberg observed that Einstein was not excessively troubled with his inability to square the notion of light quanta with the wave view of light, "nor did he even attempt to remove the inconsistency in the interpretation. He simply took the contradiction as something which would probably be understood only much later".[8] Heisenberg and Bohr developed a new way of looking at physical reality: acknowledging that the simultaneous use of different classical concepts to describe physical reality or experience leads to contradiction, they resorted to idea of complementarity: incompatible approaches may be needed to give the full picture.

The notion of complementarity cannot, of course, be taken at face value. We have been taught (by Quine) that, in principle, everything is for sale for a suit-

able price, but also that principles of (classical) logic are particularly dear. Surely, scientists are forced to reasses their empirical beliefs all the time, but the principles of reasoning, or rather the laws of logic, are only to be given up if nothing else helps. Now, we are suggested that scientists face cases where nothing else helps, at least if they respect logic. I shall start with some minor comments and then focus on what I think is at stake.

The first observation to make here is that Meheus is in the company of the friends of discovery. In classical logic, accepting contradictory premises leads to accepting anything. The logic, therefore, cannot give guidance to the search of new hypotheses – it fails as a logic of discovery. One standard objection to proposed logics of discovery has been that it is not possible to make inferences from an observation. And an equally standard objection to, e.g., Hanson's and Peirce's schemas for abductive inference, has been that they in a sense presuppose what they are supposed to explain. They presuppose, namely, that the hypothesis (or hypotheses) which are assessed with respect to their power to make the unexpected event a matter of course is already available. But in the case of the discovery of Uranus, the discovery was based on inference and reasoning. It was inferred from, on the one hand, the fact that the object appeared especially large and that it moved among the stars, and, on the other hand, the theoretical assumption that comets have both these properties.[9] Similarly, if Meheus is right, Clausius actually was also engaged in clear-cut reasoning into Carnot's theorem.

The focus here is on the reasoning process which starts long before the hypotheses surfaces. Now, although not all the details of the reasoning process are specified, it is nevertheless clear that the ideas did not emerge out of unstructured nothingness. Even stronger, reasoning from the inconsistencies was seen as necessary to arrive at a satisfactory replacement of the inconsistent theory. The reason for this is not difficult to understand. In order to restore consistency, some parts of the theory had to be rejected or modified. However, in order for this selection to be rational, it had to be based on arguments. One could object that a rational agent never needs (or must) accept an inconsistent set of beliefs. Meheus's point, however, could be that Clausius could not have taken the crucial steps in the reasoning without using both of the two inconsistent theories, those of Carnot and Joule.

6. ARGUMENT, INFERENCE AND REASONING

But let us return to the quotation from which we started, that scientists do not use any logic. I have suggested that making observations contradict theories requires some non-trivial stage setting. But it also occasions a more specific comment on the distinction between logical *arguments* on one hand and *inference* and *reasoning* on the other hand. Meheus carefully distinguishes between individual inferences and reasoning processes, where the latter are understood as series of individual inferences. Consequently, the term "premise" is ambiguous: some-

times it refers to the premises of an individual inference, sometimes to the starting point of a reasoning process. Now, individual inferences are characterized by her as inductive or deductive, depending on whether the inference is truth-preserving. This way of putting the distinction is common, and so are the further ways of characterizing inductive inferences as non-monotonic and conditional on what further premises might be added.

Meheus writes (towards the end of section 1) that she does not want to quibble over the term "logic" but rather uses it "to refer to any system that enables one, for a specific application context, to distinguish sound and unsound reasoning, and that moreover is characterized in a formally decent way." Without quibbling over words, consider the following quotation from Meheus:

> When the distinction is characterized in this way, it seems natural to assume that there exists some general criterion to classify different inferences (and, by extension, different reasoning processes) into inductive and deductive.

Deductive inference here would refer to using a deductive argument to draw a conclusion, whereas inductive inference would use some ampliative pattern to draw a conclusion. By extension, inductive inference can be thought to cover reasoning process which contain more than one individual inference. As Meheus notes, formal logic is happy to talk about deductive inferences, but hesitates to accommodate inductive inference, on account of the fact that there are no formal rules for soundness (here we should perhaps say validity) to match inferences using deductively valid argument patterns. The upshot seems to be that the soundness of inductive inferences must, if such there is, "rely on hidden non-formal reasoning".

It is clear that logic can be used to assess inferences as well as reasoning or reasoning processes. Admittedly, also, talk of deductive and inductive inference is a commonplace. However, there is a further distinction to be made between the task of logic in the sense of investigating valid arguments or argument patterns and the study of inferences and reasoning. Arguments, valid and invalid, consist of premises and a conclusion, where premises are sentences or statements or possibly proposition. Whatever the metaphysical status of premises and conclusions, a valid argument stays valid whether or not it is used by anyone to infer or reason to anything. Indeed, the premise and the conclusion of a logical argument are what they are whether or not someone happens to believe them. Inferences, in contradistinction to this, have to do with someone using some premises to form a belief, and reasoning is about a possibly more complex psychological (and some would say even social) process which leads to someone's coming to accept or believe that something is or is not the case.

"Premise", therefore, is three-ways ambiguous (at least), referring as it does to the premises of an argument, premises someone uses to form a belief (a basis of inferring that something is the case), and premises which form the starting point of a series of such inferences. Representing a piece of scientific reasoning, therefore, involves a series of interpretations. Furthermore, a logically valid pat-

tern of argument as such cannot force someone to make the corresponding inference. One may, so to speak, entertain an argument which conforms to the pattern of *modus ponens*, but refuse to accept the conclusion, if one thinks that the conclusion is false or even dubious. This, I think, is the way we understand inference using *Reductio ad Absurdum*. The formal pattern is one thing, rejecting a premise of a formally impeccable argument another thing. (I think we learned the lesson from what Achilleus said to the Tortoise).

It is obvious, as Meheus says, that in actual cases of reasoning processes, inductive and deductive inferences often intertwine. This I read to mean that when scientists and laymen come to believe that something is the case they use inferences, some of which are deductively valid (in the happy case), and others, good inductive arguments (in the happy case). However, she then writes that "it seems difficult to find examples of reasoning processes that, on the above characterization, are purely deductive". But surely there are reasoning processes which only resort to valid argument patterns. If deductively valid arguments are chained with one another (if this visual expression is allowed) the result can be represented by a deductively valid argument pattern (bearing in mind the usual strictures against equivocation). Here we could speak of deductive *reasoning* unless there were the danger of running together a logical pattern and the use of this pattern to illustrate that someone has arrived at a conclusion via a chain of valid arguments. If we choose to so talk about reasoning, we should bear in mind that this reasoning, something the reasoner does, is good reasoning, because there is a valid pattern of inference from the premises to the conclusion.

Whether and how often scientists actually engage in purely deductive reasoning in this sense is moot, but in a sense irrelevant, for it turns on whether chaining individual deductive inferences together serves a further purpose. But when inductive inferences are chained with one another (pardon the image again), the result cannot be represented by a chain of logically valid patterns of argument. With the distinction between logical argument and inference at hand, one should also be somewhat weary of the way induction and deduction mingle: "inferences that in one context clearly classify as deductive may in a different context appear to be inductive" (because conditional and non-monotonic). This of course is true, if by inference we mean someone in fact or rationally coming to believe. In the same vain, talking about sound and unsound inferences should be taken with caution: both sound (and unsound) inferences might lead to accepting something that is false.

There is a deep issue which, nevertheless, cannot be dealt with here. It might be thought that deduction is, strictly speaking, a matter of formal logical argument, whereas induction falls in the province of inference and reasoning – after all, what else does its non-monotonicity mean (this might be one way of understanding Popper's crusade against inductive logic)?

7. ACTUAL PRACTICE – WHAT AND WHY?

But perhaps this is quibbling with words. Logic can be understood narrowly as referring to the study of valid arguments in the formal sense. But there is also the wider and more traditional view in which logic covers all reasoning, ampliative and non-ampliative alike, and provides normative tools for assessing both the processes and the results of such reasoning. Logic of science, both logic of justification and logic of discovery, is best understood in this sense. So here the claim is not that new formally valid patterns are invented or discovered, but that those that already are in use are put to use in describing actual practice.

I find this aim unobjectionable, and only want to add two further considerations, one about the aims of reconstruction and the other one about the role of adaptive logic. Recall the basic pattern of objection to using logic to explicate the notion of a scientific theory, or the notion of evidential support, or the notion of explanation, not to speak of scientific discovery and other (alleged) blank spots on the map of the logic of science. The objection, since the 1950's, was that logic should be replaced by or supplemented with pragmatic analyses. This complaint was voiced again and again by the critics of logical empiricism with the result that the community of philosophers of science was split, some seeking consolation on the arms of historians and sociologists of science. But there also was an additional though often unnoticed further complaint in the claim of irrelevance of logic, viz., that whatever metatheory is offered it should be one under which the scientists themselves view their work. There is an obvious parallel in the philosophy of the social sciences concerning the concepts available for an anthropologist or sociologist in the description and explanation of an alien (often also "primitive", as in early anthropology) practice of "form of life". And indeed this discussion has found its way to philosophy of science also.

This is not the place to wander deep into this direction, but there is understanding and there is understanding. There need be no unique mode of understanding science, nor one metatheory to describe actual practice. For of course these metatheoretical notions serve a purpose which is different from that of historians and sociologists of science. For this reason, also, the tools and means of analysis could be different. It is not obvious, either, that the outcomes of these analyses should be accessible to scientists themselves, at least not without effort. So if someone comes up with a new type of logic to capture what scientists are actually doing, one should both show that scientists do use these patters of reasoning and that these patterns are illuminating. The point of course is that there can be different ways in which a logic is illuminating.

Of course, it is a legitimate question as to what scientists actually do and how they reason – and this can be conceived both as thinking and as doing. But there is a further difficulty about focusing on how scientists actually reason. If we ask them as to how they really came about their ideas, or how they really thought

these ideas were worthy of pursuit, or how they validated their results to themselves and the scientific community, we are likely to hear an account which has been tidied with the benefit of hindsight. Moreover, the way these accounts are tidied often show surprising similarities with the extant methodological and philosophical doctrines and even fashions. Thus a Victorian scientist, in England that is, would have had to pay homage to Bacon's view of scientific inquiry or to Newton's *Rules of Reasoning*, and later to explicators such as William Whewell of John Stuart Mill. The trouble is not just that the practices and ideals of scientists exhibited evolutionary and sometimes revolutionary changes, the trouble is that the metatheoretical notions, such as induction and deduction, exhibited similar changes. The inductive view is a particularly telling one. In July 1837, Charles Darwin recorded in his autobiography:

> I worked on true Baconian principles, and without any theory collected facts on a wholesale scale, more especially with respect to domesticated productions, by printed enquiries, by conversation with skilful breeders and gardeners, and by extensive reading.[10]

As Darwin scholars know, Darwin completely missed the logic of his actual practice. Bacon's philosophical and methodological views of course differed very much from the view called Baconianism.

8. PHILOSOPHICAL CONSILIENCE – AND SOME OF ITS DANGERS

I fully agree with Meheus in thinking that philosophers conceded too much to critics of logic in their wish to appreciate actual practice. But one may wonder if adaptive logic goes too far in the opposite direction, for precisely the same respect for actual practice: adaptive logic is well-behaved, but its main import is that it is not confined to pieces of reasoning which instantiate explicit formally valid patterns of inference, but capture the entire path of inquiry from initial conception to acceptance. But again the point is: all reconstructions of such paths of inquiry are just that, reconstructions.

What then is the criterion for choosing between rival reconstructions? Meheus considers but dismisses the claim that the study of dynamic reasoning patterns is "not a matter of logic, but of heuristics". Its value is in capturing the dynamics of reasoning, in adapting to inductive, deductive and abductive contexts, in accommodating discovery and not just justification. Pooling together these (and other) advantages suggests that behind all this there is the overarching desideratum of philosophical consilience. More specifically, the desideratum is that of fitting the diversity of uses of logical moves into a coherent overall pattern, integrating recognized, but separate uses or types of move into a unitary account.

This comment is inspired by the overall agreement that growth of knowledge and actual practice are too important topics to leave to historians, sociologists and psychologists. The question, still open I think, is whether extending logic to

cover inference and reasoning, the way I suggest they do not, just unifies but also obliterates distinctions we want to have. We would like to hear more of these themes, but the space does not allow but mention of them.

NOTES

1. A good example of this trend is one of the best text books on scientific reasoning, Ronald Giere's *Understanding Scientific Reasoning*. The first edition included an excellent example of how semi-formal inductive and deductive inferences must be pooled together to set up an argument for accepting or rejecting a theoretical hypothesis, given the results obtained from experiments and observations. However, the third volume is completely different: "I had reasons for dissatisfaction with this approach long before I conceived an alternative. From the students' side, representing scientific reasoning in terms of simple argument forms seemed to me not to contribute much to their understanding of the scientific context. Reconstructing the reasoning as an explicit argument was often a burdensome, and frequently mechanical, exercise". In the third edition the introductory chapters with deductive arguments have been completely eliminated – a change which "has made possible considerable simplification and streamlining of the whole text." Ronald N. Giere, *Understanding Scientific Reasoning*. New York: Holt, Rinehart and Winston, Third edition, 1991, Preface, v. (First edition, 1979).
2. Meheus actually goes further than this by suggesting that her analysis of the reasoning of Clausius is not a rational reconstruction in Lakatos's sense but "an attempt to reconstruct, on the basis of the original texts, as accurately as possible Clausius's reasoning process. Joke Meheus, "Adaptive Logic in Scientific Discovery: the Case of Clausius", *Logique et Analyse*, 143-144, p. 360.
3. Imre Lakatos, "Falsification and the Methodology of Scientific Research Programmes". In I. Lakatos and A. Musgrave (eds.), *Criticism and the Growth of Knowledge*. Cambridge: Cambridge University Press 1970, p. 91.
4. For the distinction between empirical and conceptual problems, see Larry Laudan, *Progress and Its Problems; Towards a Theory of Scientific Growth*. London: Routledge and Kegan Paul 1977.
5. Peter Clark gives a detailed Lakatosian account of the rise of thermodynamics. As Meheus (note 6) observes, Clark's is a reconstruction which builds on the notion that Clausius did not reason from inconsistency but first resolved it. This indeed is the case. However, she also suggests that this resolution was done "on the basis of non-logical grounds". The crucial issue here is what non-logical grounds here means: identifying the inconsistency as an inconsistency involves using patterns of inference which can be explicated by help of formal argument schemas. But applying logic to identify inferences seems empty without reference to beliefs which have content. On the other hand beliefs without logical structure are blind. I shall return to this question below. See Peter Clark, "Atomism versus thermodynamics", in: Colin Howson (ed.), Method and Appraisal in the Physical Sciences. The Critical Background to Modern Science 1800-1905. Cambridge: Cambridge University Press 1976, pp. 41-105.
6. There is also the further complication which is particularly conspicuous in the case of Thomas Kuhn's views of paradigms and its later formal incarnation in the form of the so-called structuralist proposal. The latter can be seen as one explication of the former, and its early version was in a sense sanctioned by Kuhn in T.S. Kuhn, "Theory Change as Structure Change: Comments on the Sneed Formalism", *Erkenntnis*, 10, pp. 179-199. The complication is that truth falls out of the picture altogether. Part of this idea is that the fundamental laws (e.g. the Newtonian laws) do not describe the reality directly but only indirectly, through a series of steps through special laws, initial conditions, etc., which describe a particular system. For the structuralist view, see Wolfgang Balzer and Carlos Ulises Moulines (eds.), *Structuralist Theory of Science: Focal Issues, New Results*. Berlin and New York: Walter de Gruyter 1996.
7. Joke Meheus, *ibid*, pp. 359-389.

8. Werner Heisenberg, *Physics and Philosophy*, Penguin Books, 1989, p. 21.
9. Peter Achinstein, for instance, has objected to the view that discovery is an irrational or arational affair, up to imagination, inspiration or unfounded conjecture. Achinstein suggests that Galileo literally inferred that the moon has craters on the basis of what he saw through the telescope. Moreover, this inference was based on evidential reasons. See Peter Achinstein, "Discovery and Rule-Books", in Thomas Nickles (ed.), *Scientific Discovery, Logic and Rationality*. Dordrecht: Holland/Boston: USA, D. Reidel, 1980, pp. 117-137.
10. Charles Darwin, C. *More Letters of Charles Darwin*. (Two volumes). Edited by F. Darwin and A.C. Seward. New York: Appleton & Co., 1902-3.

Department of Mathematics, Statistics and Philosophy
University of Tampere
33014 University of Tampere
matti.sintonen@helsinki.fi

WOLFGANG SPOHN

LAWS ARE PERSISTENT INDUCTIVE SCHEMES

1. INTRODUCTION [1]

Laws are true lawlike sentences. But what is lawlikeness? Much effort went into investigating the issue, but the richer the concert of opinions became, the more apparent their deficiencies became too, and with it the profound importance of the issue for epistemology and philosophy of science.

The most widely agreed prime features are that laws, in contrast to accidental generalizations, support counterfactuals, have explanatory power, and are projectible from, or confirmed by, their instances. These characteristics have long been recognized. However, the three topics they refer to – counterfactuals, explanation, and induction – were little elaborated in the beginning and are strongly contested nowadays. Moreover, the interrelations between these subjects were quite obscure. Hence, these features did not point to a clear view of lawlikeness, either. In this paper, I try to advance the issue. Let me start with three straight decisions.

The first decision takes a stance on the priority of the prime features. I am convinced that it is the inductive behavior associated with laws which is the most basic one, and that it somehow entails the other prime features. I cannot justify this stance in a few lines. Suffice it to say that my study of causation (1983) led me from Lewis' (1973) theory of counterfactuals over Gärdenfors' epistemic account of counterfactuals (cf., e.g., Gärdenfors 1981) ever deeper into the theory of induction where I finally thought I had reached firm ground. Clearly, this is not an unusual intellectual movement. In Spohn (1991), I explained my view on the relation of induction to causation, and thus to explanation. However, I did not return to counterfactuals, not because I thought I had inadequate means for analyzing them, but because I felt that this subject is overlaid by many linguistic intricacies that are hard and perhaps not really important to account for. My decision finds recent support, for instance, in Lange (2000) who starts investigating the relation between laws and counterfactuals and also arrives at induction as the most basic issue.[2]

The second decision concerns the relation between laws and their prime features. When inquiring into lawlikeness, the idea often was to search for something which *allows* us to use laws in induction, explanation and counterfactuals in the way we do. That is, given that induction is really the most basic aspect,

F. Stadler (ed.), Induction and Deduction in the Sciences, 135–150.
© 2004 *Kluwer Academic Publishers.*

lawlikeness should be something that *justifies* the role of laws in induction. This idea issued in perplexity; no good candidate could be found providing this justification.

There is an alternative idea, namely that lawlikeness is *nothing but* the role of laws in induction. In view of the history of inductive scepticism from Hume to Goodman – which made us despair of finding a deeper justification of induction and taught us rather to describe our inductive behavior and to inquire what is rational about it while being aware that this inquiry may produce only partial justification – this idea seems to be the wiser one. I do not mean to suggest that the lessons of inductive scepticism have been neglected.[3] But it is important to be fully aware of these lessons, and hence I shall pursue here the second idea and foreswear the search for deeper justifications. We shall see that we can still say quite a lot about rational induction.

We are thus to study the inductive properties of laws. This presupposes some account of induction or confirmation within which to carry out the study. This is what my third decision is about. I think that on this matter philosophy of science went entirely wrong in the last 25 years. Bayesianism was always strong, and rightly so. In the 1950's and 60's, much effort also went into the elaboration of a qualitative confirmation theory. However, this project was abandoned in the 70's. The main reason was certainly that the efforts were not successful at all. Niiniluoto (1972) gives an excellent survey that displays the incoherencies of the various attempts. An additional reason may have been the rise and success of the theory of counterfactuals, which answered many problems in philosophy of science (though not problems of induction) and thus attracted a lot of the motivation originally directed to an account of induction.

In any case, the effect was that Bayesianism was more or less the only remaining well-elaborated alternative. This hindered progress, because deterministic laws and probability do not fit together well. Deterministic laws are not simply the limiting case of probabilistic laws, just as deterministic causation is not the limiting case of probabilistic causation.[4] We find a parallel in the disparity between belief, or acceptance-as-true, and subjective probability, which was highlighted by the lottery paradox and has, I think, as yet not found a convincing reconciliation. My conclusion is, though I have hardly argued for it, that Bayesianism is of little help in advancing the issue of lawlikeness.

Philosophical logic was very active since around 1975 in producing alternatives, though not under the labels 'induction' or 'confirmation'. However, these activities were hardly recognized in philosophy of science. Instead, they radiated to AI where they were rather successful. It is precisely in this area where we shall find help. Let me explain.

What should we expect an account of induction or inductive schemes to achieve? I take the view (cf. Spohn 2000a) that it is equivalent to a theory of belief revision or, more generally, to an account of the dynamics of doxastic states. This is why the topic is so inexhaustible. Everybody, from the neurophysiologist

to the historian of ideas, can contribute to it, and one can deal with it from a descriptive as well as from a normative perspective.

Philosophers, I assume, would like to come up with a very general normative account. Bayesianism provides such an account that is almost complete. There, rational doxastic states are described by probability measures, and their rational dynamics is described by various conditionalization rules. As mentioned, however, in order to connect up with deterministic laws, we should proceed with an account of doxastic states which represents plain belief or acceptance-as-true. Doxastic logic is sufficient for the statics, but it does not provide any dynamics. Probability < 1 cannot represent belief, because it does not license the inference from the beliefs in two conjuncts to the belief in their conjunction. Probability 1 cannot do it either, because we would like to be able to update with respect to information previously disbelieved, because disbelieved propositions would have probability 0 according to this approach, and because Bayesian dynamics does not provide an account of conditionalization with respect to null propositions (that is why I called Bayesianism almost complete). Hence, Bayesianism is unhelpful. Belief revision theory (cf., e.g., Gärdenfors 1988) was devised to fill the gap. Unfortunately, the dynamics it provides turned out to be incomplete as well (cf. Spohn 1988, sect. 3). There have been several attempts to plug the holes (cf., e.g., Nayak 1994 and Halpern 2001), but I am still convinced that ranking theory, proposed in Spohn (1983, sect. 5.3, and 1988), though under a different name, offers the most adequate account for a full dynamics of plain belief.[5]

In any case, this is my third decision: to carry out my study of the inductive behavior of laws strictly in terms of the theory of ranking functions. This framework may be unfamiliar, but the study will not be difficult, since ranking theory is a very simple theory. Still, my constructive attempts will leave little space for broader discussion. The hope is that my account appears illuminating even without extensive comparative discussion.

The plan of the paper is now almost obvious. In section 2, I shall introduce the theory of ranking functions as far as needed. Section 3 explicates lawlikeness, i.e., the difference between laws and accidential generalizations insofar as it can be expressed in ranking terms. At least a brief comparative discussion is certainly called for; this is given in section 4. Since belief in a law is analyzed not as a belief in a regularity or some more sophisticated proposition, but rather as a certain inductive attitude, the immediate question arises how a law, i.e., such an inductive attitude, can be confirmed in turn. This crucial question is addressed in the final section 5.

I do not deny in this paper, but only neglect that laws have important metaphysical aspects as well. I have been less negligent in Spohn (1993), where I tried to understand causal laws as objectifications of inductive schemes, and in Spohn (1997), where I discussed both aspects of reduction sentences, the laws associated with disposition predicates. But if I am right, lawlikeness is an epistemological criterion, and this is why this paper entirely focuses on the epistemological aspects of laws.

2. RANKING FUNCTIONS

Let us start with a set W of possible worlds, small rather than large worlds, as we shall see soon. Each subset of W is a truth condition or *proposition*. I assume propositions to be the objects of doxastic attitudes. Thus, I take these attitudes to be intensional. We know well that this is problematic, and we scarcely know what to do about the problem. Hence, my assumption is just an act of front alignment.

The assumption also entails that we need not distinguish between propositions and sentences expressing them. Hence, I shall often use first-order sentences to represent or denote propositions and shall not distinguish between logically equivalent sentences, since they express the same proposition.

That is all we need to introduce our basic notion: κ is a *ranking function* (for W) iff κ is a function from W into $\mathbf{N}$ (the set of non-negative integers) such that $\kappa(w) = 0$ for some $w \in W$. For each proposition $A \subseteq W$ the *rank* $\kappa(A)$ of A is defined by $\kappa(A) = \min \{\kappa(w) \mid w \in A\}$ and $\kappa(\varnothing) = \infty$. For $A, B \subseteq W$ the (*conditional*) *rank* $\kappa(B \mid A)$ of B *given* A is defined by $\kappa(B \mid A) = \kappa(A \cap B) - \kappa(A)$. Since singletons of worlds are propositions as well, the point and the set function are interdefinable. The point function is simpler, but auxiliary; the set function is the one to be interpreted as a doxastic state.

Indeed, ranks are best interpreted as *grades of disbelief*. $\kappa(A) = 0$ says that A is not disbelieved at all. It does not say that A is believed; this is rather expressed by $\kappa(\overline{A}) > 0$,[6] i.e., that non-A is disbelieved (to some degree).[7] The clause that $\kappa(w) = 0$ for some $w \in W$ is thus a *consistency* requirement. It guarantees that at least some proposition, and in particular W itself, is not disbelieved. This entails the *law of negation*: for each $A \subseteq W$, either $\kappa(A) = 0$ or $\kappa(\overline{A}) = 0$ or both.

The set $C_\kappa = \{w \mid \kappa(w) = 0\}$ is called the *core* of κ (or of the doxastic state represented by κ). C_κ is the strongest proposition believed (to be true) in κ. Indeed, a proposition is believed in κ if and only if it is a superset of C_κ. Hence, the set of beliefs is *deductively closed* according to this representation.[8]

There are two laws for the distribution of grades of disbelief. The *law of conjunction*: $\kappa(A \cap B) = \kappa(A) + \kappa(B \mid A)$, i.e., the grade of disbelief in A and the grade of disbelief in B given A add up to the grade of disbelief in A-and-B. And the *law of disjunction*: $\kappa(A \cup B) = \min\{\kappa(A), \kappa(B)\}$, i.e., the grade of disbelief in a disjunction is the minimum of the grades of the disjuncts. The latter is again only a consistency requirement, though a conditional one; if that law would not hold the inconsistency could arise that both $\kappa(A \mid A \cup B), \kappa(B \mid A \cup B) > 0$, i.e., that both A and B are disbelieved given A-or-B.

According to the above definition, the law of disjunction indeed extends to disjunctions of arbitrary cardinality. I find this reasonable, since an inconsistency is to be avoided in any case, be it finitely or infinitely generated. Note that this entails that each countable set of ranks has a minimum, and thus that the range of a ranking function is well-ordered. Hence, the range $\mathbf{N}$ is a natural choice.[9]

However, here we better avoid all complexities involved in infinity. Therefore, I shall outright assume that we are dealing only with finitely many worlds and hence only with finitely many propositions. This entails that each world in W (or the set of its distinctive features) is finite in turn. Hence, as announced, they are small worlds. One may think that this is a strange start for an investigation of natural laws. However, an analysis of lawlikeness should work also under this finiteness assumption. After all, our world seems both to have laws and to be finite. Generalizing my observations below to the infinite case is urgent, but would require a separate paper.

There is no need here to develop ranking theory extensively. A general remark may be more helpful: ranking theory works in almost perfect parallel to probability theory. Take any probabilistic theorem, replace probabilities by ranks, the sum of probabilities by the minimum of ranks, the product of probabilities by the sum of ranks, and the quotient of probabilities by the difference of ranks, and you are almost guaranteed to arrive at a ranking theorem. For instance, you thus get a ranking version of Bayes' theorem. Or you can develop the whole theory of Bayesian nets in ranking terms. And so on. The general reason is that one can roughly interpret ranks as the orders of magnitude of (infinitesimal) probabilities.

The parallel extends to the laws of doxastic change, i.e., to rules of conditionalization. Thus, it is at least plausible that ranking theory provides a complete dynamics of doxastic states (as may be shown in detail; cf. Spohn, 1988, sect. 5). It is due to this fact that ranking functions are fully entitled to be referred to by my loose talk of inductive schemes in the title of this paper.

It is still annoying, perhaps, that belief is not characterized in a positive way. But there is remedy: β is the *belief function* associated with κ (and thus a belief function) iff β is the function assigning integers to propositions such that $\beta(A) = \kappa(\overline{A}) - \kappa(A)$ for each $A \subseteq W$. Similarly, $\beta(B \mid A) = \kappa(\overline{B} \mid A) - \kappa(B \mid A)$. Recall that at least one of the terms $\kappa(\overline{A})$ and $\kappa(A)$ must be 0. Hence, $\beta(A) > 0, < 0$, or $= 0$ iff, respectively, A is believed, disbelieved, or neither; and A is the more strongly believed, the larger $\beta(A)$. Thus, belief functions may appear to be more natural. But their formal behavior is more awkward. Therefore, I shall use both notions.

Above, I claimed that a full dynamics of belief is tantamount to an account of induction and confirmation. So, what is confirmation with respect to ranking functions? The same as elsewhere, namely *positive relevance*: *A confirms* or is a *reason for B* relative to κ iff $\beta(B \mid A) > \beta(B \mid \overline{A})$, i.e., iff $\kappa(\overline{B} \mid A) > \kappa(\overline{B} \mid \overline{A})$ or $\kappa(B \mid A) < \kappa(B \mid \overline{A})$ or both.[10]

There is an issue here whether the condition should require $\beta(B \mid A) > \beta(B)$ or only $\beta(B \mid A) > \beta(B \mid \overline{A})$, as stated. In the corresponding probabilistic case, the two conditions are equivalent if all three terms are defined, but the first condition is a bit more general, since it may be defined while the second is not. That is why the first is often preferred. In the ranking case, however, all three terms are always defined, and the second condition may be satisfied while the first is not. In

that case, the second condition on which my definition is based seems to be more adequate.[11]

A final point that will prove relevant later on: Ranking functions can be mixed, just as probability measures can. For instance, if κ_1 and κ_2 are two ranking functions for W and if κ^* is defined by

$$\kappa^*(A) = \min\{\kappa_1(A), \kappa_2(A) + n\} \text{ for some } n \in \mathbf{N} \text{ and all } A \subseteq W,$$

then κ^* is again a ranking function for W. Or more generally, if K is a set of ranking functions for W and ρ a ranking function for K, then κ^* defined by

$$\kappa^*(A) = \min\{\kappa(A) + \rho(\kappa) \mid \kappa \in K\} \text{ for all } A \subseteq W$$

is a ranking function for W. The function κ^* may be called the *mixture* of K by ρ.

This is all the material we shall need. I hope that the power and beauty of ranking theory is apparent already from this brief introduction. I have not argued here that if one wants to state a full dynamics of plain belief or acceptance-as-true, one must buy into ranking theory. I did so in Spohn (1988, sect. 3). Even that argument may not be entirely conclusive. However, I guess the space of choices is small, and I would be very surprised if a simpler choice than ranking theory were to be available.

Be this as it may, let us finally turn to our proper topic, the epistemology of laws.

3. LAWS

Let me start with a simple formal observation. Given some ranking function κ, to believe $A \wedge B$ means that $C_\kappa \subseteq A \cap B$, i.e., $\kappa(\neg A \vee \neg B) > 0$, i.e., $\min\{\kappa(\neg A), \kappa(\neg B)\} > 0$. This, however, can be implemented in many different ways. In particular, it leaves open how $\kappa(\neg A \vee \neg B)$ relates to $\kappa(\neg A)$ and $\kappa(\neg B)$ and whether and how A and B depend on each other. Hence, if you start with believing $A \wedge B$, but now learn that $\neg A$ obtains, you may, or may not, continue to believe B, depending on the value of $\kappa(\neg B \mid \neg A)$.

Basically the same point applies to believing a universal generalization. This, I propose, is the clue to understanding laws. Let us take $G = \wedge x(Px \rightarrow Qx)$ as our prototypical generalization ($\rightarrow$ always denotes truth-functional material implication). I have already simplified things by assuming the worlds in W to be finite. This entails that the quantifier in G ranges over some finite domain D. For $a \in D$, let G_a be the instantiation of G by a, i.e., $G_a = Pa \rightarrow Qa$. Now to believe G in κ means that $C_\kappa \subseteq G$, i.e., $\kappa(\neg G) > 0$, i.e., $\min\{\kappa(\neg G_a) \mid a \in D\} > 0$. Thus, the generalization is believed as strongly as the weakest instantiation.[12]

Let us assume, moreover, that this is the only belief in κ, i.e., that $C_\kappa = G$; thus, no further beliefs interfere. This entails in particular that $\kappa(Pa \wedge Qa) =$

$\kappa(\neg Pa \wedge Qa) = \kappa(\neg Pa \wedge \neg Qa) = 0 < \kappa(Pa \wedge \neg Qa)$ for each $a \in D$ and hence that $\kappa(\neg Qa \mid Pa) > 0$, i.e., that Pa is positively relevant for Qa. In other words, under this assumption the belief in the material implication $Pa \rightarrow Qa$ is equivalent to the positive relevance of Pa for Qa.

Again, the belief in G can be realized in many different ways. Let me focus for a while on two particular ways, which I call the 'persistent' and the 'shaky' attitude. If you learn about positive instances, G_a, G_b, etc., you do not change your beliefs according to κ, since you expected them to be positive anyway.[13] The crucial difference emerges when we look at how you respond to negative instances, $\neg G_a$, $\neg G_b$, etc., according to the various attitudes:

If you have the *persistent* attitude,[14] your belief in further instantiations is unaffected by negative instances, i.e., $\kappa(\neg G_b) = \kappa(\neg G_b \mid \neg G_a)$ $(b \neq a)$, and indeed $\kappa(\neg G_b) = \kappa(\neg G_b \mid \neg G_{a_1} \wedge \ldots \wedge \neg G_{a_n})$ for any $n \in \mathbf{N}$ $(b \neq a_1,\ldots,a_n)$. If, by contrast, you have the *shaky* attitude, your belief in further instantiations is destroyed by a negative instance, i.e., $\kappa(\neg G_b \mid \neg G_a) = 0$ and, a fortiori, $\kappa(\neg G_{\neq a} \mid \neg G_a) = 0$.[15]

The difference is, I find, characteristic of the distinction between lawlike and accidental generalizations. Let us look at two famous examples. First the coins:

(1) All German coins are round.

(2) All of the coins in my pocket today are made of silver.

It seems intuitively clear to me that we have the persistent attitude towards (1) and the shaky attitude towards (2). If we come across a cornered German coin, we wonder what might have happened to it, but our confidence that the next coin will be round again is not shattered. If, however, I find a copper coin in my pocket, my expectations concerning the further coins simply collapse; if (2) has proved wrong in one case, it may prove wrong in any case.

Or look at the metal cubes, which are often thought to be the toughest example:

(3) All solid uranium cubes are smaller than one cubic mile.

(4) All solid gold cubes are smaller than one cubic mile.

What I said about (1) and (2) applies here as well, I find. If we bump into a gold cube this large, we are surprised – and start thinking there might well be further ones. If we stumble upon an uranium cube of this size, we are surprised again. But we find our reasons for thinking that such a cube cannot exist unafflicted and will instead start investigating this extraordinary case (if it obtains for long enough).

The example of Bode's law mentioned by Köhler (2003) fits perfectly. As Köhler explains, the law appeared to be entirely accidental for a long time. This finds expression in a shaky attitude. If astronomists had discovered a new planet

in our solar system, or even another solar system with localizable planets, which does not obey Bode's law, we would have abandoned it, i.e., we would have lost all confidence to apply it to new cases. But, as Köhler goes on to explain, it has nowadays acquired the status of a genuine law. If such an exception were discovered today, we would seek for possible explanations of this exception, but still apply the law to new cases, i.e., we would display a persistent attitude.[16]

As far as I see, the difference between the shaky and the persistent attitude applies as well to the other examples prominent in the literature.[17] However, my wording is certainly more determined than my thinking. According to my survey, intuitions are often undecided. There may be mixed attitudes, and somehow the attitude seems to depend on how one came to believe in the regularity; there may be different settings for one and the same generalization. At the moment, though, I am concerned with carving out what appears to me to be the basic difference, and, therefore, I am painting black and white.

However, it is easy to refine the distinction between the persistent and the shaky attitude and to gain a systematic overview over the range of options provided by ranking theory. Let me briefly explain this point before further discussing my explication.

How many ways are there to believe the generalization G, i.e., to realize $\kappa(\neg G) > 0$? A natural and strongly simplifying assumption is

Symmetry: For all $a_1,...,a_n,b_1,...,b_n \in D$

$$\kappa(\neg G_{a_1} \wedge ... \wedge \neg G_{a_n}) = \kappa(\neg G_{b_1} \wedge ... \wedge \neg G_{b_n}).$$

In obvious analogy to inductive logic, symmetry says that the disbelief in violations of a generalization depends on their number, but not on the particular instances. For $n = 1$ symmetry entails that there is some $r > 0$ such that for all $a \in D$ $\kappa(\neg G_a) = \kappa(\neg G) = r$. More generally, symmetry entails, as is easy to see, that there is some function c from $\mathbf{N}$ to $\mathbf{N}$ such that for any $n+1$ different $a_1,...,a_n,b \in D$ the equality $\kappa(\neg G_b \mid \neg G_{a_1} \wedge ... \wedge \neg G_{a_n}) = c(n)$ holds, where $c(0) = r$. Indeed, all ranks of all Boolean combinations of the G_a are uniquely determined by the function c.

Another plausible assumption familiar from inductive logic is

Non-negative instantial relevance: For all $a_1,...,a_n,a_{n+1},b \in D$

$$\kappa(\neg G_b \mid \neg G_{a_1} \wedge ... \wedge \neg G_{a_n}) \geq \kappa(\neg G_b \mid \neg G_{a_1} \wedge ... \wedge \neg G_{a_n} \wedge \neg G_{a_{n+1}}).$$

This is tantamount to the function c being non-increasing.

Given the two assumptions there remain not so many ways to believe G; any non-increasing function c with $c(0) = r$ stands for one such way. Hence, the persistent attitude characterized by $c(n) = r$ for all n stands for one extreme, whereas the shaky attitude for which $c(n) = 0$ for $n \geq 1$ stands for the other. So, one may

think about whether any ways in between fit the examples better than the extreme ones.

Still, the consideration shows that the two attitudes I have taken to distinguish accidental and lawlike generalizations are suited best for marking the spectrum of possible attitudes. Hence, my discussion will proceed in terms of these two extremes.

4. DISCUSSION

The examples have suggested that we treat a universal generalization G as lawlike if we have the persistent attitude towards it, and as accidental if we have the shaky attitude towards it. Hence, the difference does not lie in the propositional content, it lies only in our inductive attitude towards the generalization or, rather, its instantiations.

This account of lawlikeness is perhaps closest to the old idea that laws are not general statements, but rather inference rules or licenses. This idea goes back at least to Ramsey (1929)[18] who states very clearly: "Many sentences express cognitive attitudes without being propositions; and the difference between saying yes or no to them is not the difference between saying yes or no to a proposition" (pp. 135f.). And "... laws are not either" [namely propositions] (p. 150). Rather: "The general belief consists in *(a)* A general enunciation, *(b)* A habit of singular belief" (p. 136). Clearly, one must not read too much of the present context into Ramsey's writings; he was occupied with the problems of his days, for instance, with rejecting Russell's acknowledgment of general facts and with thus finding a different interpretation for general sentences.[19] Still, his attempts to construct a pragmatic theory of belief, which are also reflected in my brief quotations, are surprisingly modern. There does not appear to be a 70 years gap when Lange says in his much more elaborate book (2000) that "the root commitment that we undertake when believing in a law involves the belief that a given inference rule possesses certain objective properties, such as reliability" (p. 189).[20]

From a purely logical point of view, however, it was always difficult to see the difference between accepting $\wedge x(Px \rightarrow Qx)$ as an axiom and accepting the inference rule "for any a, infer Qa from Pa". The only difference is that the rule is logically weaker; the rule is made admissible by the axiom, but the axiom cannot be inferred with the help of the rule. What else beside the unproductive logical point could be meant by the slogan "laws are inference rules" was always hard to explain. Clearly, there is also no firm difference concerning confirmation, falsification, etc.; if a general sentence can be falsified by a counter-instance, then a rule can be shown to be invalid by a case in which it leads from truth to falsity.

Still, one might say that the inference-license perspective puts more emphasis on what to do in the single case. And this emphasis is not mere rhetorics. It is reflected, I think, in my central notion of persistence, and thus finds a precise

induction-theoretic basis. It should be clear here that ranking functions *are* (possibly very complex) inference rules. If inference is not understood deductively, but more liberally as confirmation or being a reason, then ranking functions may obviously be interpreted as defeasible inference rules that are believed to be reliable, even if they are not universally valid.

Hence, the mark of laws is not their universality, which breaks down with one counter-instance, but rather their operation in each single case, which is not impaired by exceptions. Here, my account meets with Cartwright (1989) and her continuous forceful efforts to explain that what we have to attend to are the capacities and their cooperation taking effect in the single case. Her objective capacities or powers thus correspond to my subjective reasons as embodied in a ranking function. And, as I tried to argue in (1993) and (2000b), the gap between the subjective and the objective and between reasons and causes is not unbridgeable.

So much for some important agreements. The most obvious disagreement is with Popper, of course. Given how much we have learned from Popper about philosophy of science, my account is really ironic, since it concludes in a way that it is the mark of laws that they are *not* falsifiable by negative instances; it is only the accidental generalizations that are so falsifiable. Of course, the idea that the belief in laws is not given up so easily is familiar at least since Kuhn's days, and even Popper (1934, ch. IV, §22) insisted from the outset that falsifications of laws proceed by more specialized counter-laws rather than simply by counter-instances, but I cannot recall having seen the point being stripped down to its induction-theoretic bones.

What I have said so far may provoke a confusion that I should hurry up to clarify. The persistent attitude towards $G = \wedge x(Px \rightarrow Qx)$ is characterized, I said, by the *independence* of the instantiations; experience of one instance does not affect belief about the others. In this way, belief about an instance G_b, i.e., the positive relevance of Pb for Qb, is persistent. But did we not learn that one mark of lawlikeness is *enumerative induction*, i.e., the confirmation of the law by positive instances? Surely, enumerative induction outright contradicts the independence I claim.

Herein lies a subtle confusion. Belief in a law is more than belief in a proposition, it is a certain doxastic attitude, and that attitude as such is characterized by the independence in question. If I would have just this attitude, just this belief in a law, my κ would exhibit this independence. Enumerative induction, by contrast, is not about what the belief in a law *is*, but about how we may acquire or confirm this belief. The two inductive attitudes involved may be easily confused, but the confusion cannot be identified as long as one thinks belief in a law is just belief in a proposition.

However, what could it mean then to confirm a law if it does not mean to confirm a proposition? Indeed, my definition of confirmation in section 2 applies only to the latter, and to talk of the confirmation of laws, i.e., of a second-order inductive attitude towards a first-order inductive attitude, is at best metaphorical

so far. Enumerative induction or falsificationism do not seem to make sense within this setting. This is a crucial point, and I shall devote the final section to it where I shall make a proposal for reconstructing enumerative induction and the falsification of laws. But so far I am concerned only with the attitude in which the belief in a law itself consists.

Is my explanation of lawlikeness a deep one? No, it is just as plain as, for instance, that of the counterfactual theorist who says that lawlikeness *is* support of counterfactuals or that a law *is* a universally quantified subjunctive conditional. Analysis has to start somewhere, and it acquires depth only by showing how to explain other features of laws by the basic ones. That is a task I cannot pursue here.[21] But I would like to insist that, as a starting point, the present analysis is to be preferred. There are good reasons for feeling uneasy about starting with subjunctives or a similarity relation between worlds. By contrast, ranking theory is a very plain theory with a very obvious interpretation.

The only doubt one may have about my starting point may concern its sufficiency as a basis of analysis. In particular, one may feel that the crucial property of laws is one which justifies the inductive attitude I have described, say, some kind of material or causal necessity. It may be. But I am sceptical and refer to my second decision in section 1.

This does not mean that I have to sink into subjectivism, that I am bound to say that it is merely a matter of one's inductive taste what one takes to be a law. There may be objectivizations and rationalizations for our beliefs in laws. I do not intend to start speculating about this, but one very general rationalization is quite obvious. It is of vital importance to us to have persistent attitudes to a substantial extent. Something is almost always going wrong with our generalizations, and if we always had the shaky attitude, our inductions and expectations would break down dramatically and we could not go on living.

5. ON THE CONFIRMATION OF LAWS

Above I have emphasized that one must not confuse the inductive or confirmatory attitude embodied in laws with the confirmation of the laws themselves. What the latter could be was, however, quite unclear, in contrast to the confirmation of propositions, which was already well handled by ranking functions. What can we do about this?

Fortunately, there is clear precedent in the literature. If one is aware of the close similarity between probability and ranking theory, one could notice that a law as I conceived it is nothing but a sequence of independent, identically distributed random variables translated into ranking terms. And, if one is familiar with the work of de Finetti (1937), one sees that he addresses exactly our problem. In his famous theorems, de Finetti showed that there is a one-one correspondence between symmetric probability measures for an infinite sequence of random variables and mixtures of Bernoulli measures according to which the

variables are independent and identically distributed, and that the mixture concentrates more and more on a single Bernoulli measure as evidence accumulates. He thus showed to the objectivist that subjective symmetric measures provide everything he wants, i.e., beliefs about statistical hypotheses that converge toward the true one with increasing evidence.

De Finetti's issue between objectivism and subjectivism is not my concern. Ranking functions are thoroughly epistemological and have as such no objective interpretation.[22] Still, we can immediately extract an account of the confirmation of laws from de Finetti's theory. Despite its artificial and formalistic appearance, the basic construction is, I find, highly illuminating.

Let us start with n mutually exclusive and jointly exhaustive properties or predicates $Q_1, \ldots, Q_n$ (these are Carnap's Q-predicates). For each $i \leq n$ we have the elementary law $G^i = \neg\forall x Q_i x = \forall x \neg Q_i x$. For any proposition $A \subseteq W$ we may now count how often the law G^i is violated if A obtains; this is done by the function $v(A, i) = \text{card}\{a \in D \mid A \subseteq Q_i a\}$.[23] So, if we define the ranking function κ^i for W by $\kappa^i(A) = v(A, i)$, κ^i precisely represents the belief in the law G^i. Without any evidence, though, we do not believe in any law G^i. Our attitude towards the laws is rather represented by the ranking function ρ_0 for which $\rho_0(\kappa^i) = 0$ for each $i = 1,\ldots,n$. Hence, our doxastic attitude towards the propositions $A \subseteq W$ is represented by the mixture κ_0 of the κ^i with respect to ρ_0, as defined by

$$\kappa_0(A) = \min_{i \leq n} \kappa^i(A) + \rho_0(\kappa^i) = \min_{i \leq n} v(A, i).$$

Now, how does this attitude change by experience? Via conditionalization, as always. But let us describe this in detail. Let $\mathbf{r} = \langle r_1,\ldots,r_n\rangle$ stand for any sequence of n non-negative integers, and let $r = r_1 + \ldots + r_n$. Define next $E(\mathbf{r})$ to be the proposition (evidence) that among the first r objects precisely r_i instantiate Q_i ($i = 1,\ldots,n$); the order of instantiation is irrelevant. Clearly, $\kappa_0(E(\mathbf{r})) = \min r_i$. Let B range over propositions about the remaining objects and not the first r ones, and let $\kappa_{\mathbf{r}}$ be the ranking function that we have for those propositions after receiving evidence $E(\mathbf{r})$. Then we have:

$$\begin{aligned}\kappa_{\mathbf{r}}(B) &= \kappa_0(B \mid E(\mathbf{r})) = \kappa_0(B \cap E(\mathbf{r})) - \kappa_0(E(\mathbf{r})) \\ &= \min_{i \leq n} (v(B, i) + r_i) - \min_{i \leq n} r_i = \min_{i \leq n} (\kappa^i(B) + (r_i - \min_{i \leq n} r_i)).\end{aligned}$$

That is, if we define $\rho_{\mathbf{r}}$ by $\rho_{\mathbf{r}}(\kappa^i) = r_i - \min_{i \leq n} r_i$, we have

$$\kappa_{\mathbf{r}}(B) = \min_{i \leq n} (\kappa^i(B) + \rho_{\mathbf{r}}(\kappa^i)).$$

Hence, $\kappa_{\mathbf{r}}$ is the mixture of the κ^i with respect to $\rho_{\mathbf{r}}$. So, the evidence $E(\mathbf{r})$ makes us change our attitude towards the laws from ρ_0 to $\rho_{\mathbf{r}}$, and $\rho_{\mathbf{r}}$ represents the de-

grees to which the various laws have been confirmed or rather disconfirmed. If $\rho_{\mathbf{r}}(\kappa^j) > 0$, we might say that κ^j is falsified, but note that falsification is never conclusive in this construction.

The definition of $\rho_{\mathbf{r}}$ also shows what enumerative induction comes to in our context. The more instances of Q_j, i.e., counter-instances to G^j we find, the higher $\rho_{\mathbf{r}}(\kappa^j)$, i.e., the stronger the disbelief in κ^j. Or a bit more formally: Suppose we have already received evidence $E(\mathbf{r})$ and now discover a new instance a_{r+1} of Q_j. Then our total evidence is $E(\mathbf{s})$, where $\mathbf{s} = \langle r_1,..., r_{j-1}, r_{j+1}, r_{j+1},...,r_n\rangle$. The relation between $\rho_{\mathbf{r}}$ and $\rho_{\mathbf{s}}$ is then immediately calculated: If $r_j \geq r_i$ for at least one $i \neq j$, i.e., if some other hypothesis is at least as good as G^j, then $\rho_{\mathbf{s}}(\kappa^j) = \rho_{\mathbf{r}}(\kappa^j) + 1$ and $\rho_{\mathbf{s}}(\kappa^i) = \rho_{\mathbf{r}}(\kappa^i)$ for $i \neq j$; i.e., the disbelief in κ^j increases. And if $r_j < r_i$ for all $i \neq j$, i.e., if G^j was so far *the* least violated hypothesis, then $\rho_{\mathbf{s}}(\kappa^j) = \rho_{\mathbf{r}}(\kappa^j) = 0$ and $\rho_{\mathbf{s}}(\kappa^i) = \rho_{\mathbf{r}}(\kappa^i) - 1$ for $i \neq j$; i.e., the other hypotheses become less disbelieved or possibly no longer disbelieved at all, and κ^j is thus possibly no longer believed. Either way, the negative instance $Q_j a_{r+1}$ is negatively relevant to κ^j and hence positively relevant to the complement $\{\kappa^i \mid i \neq j\}$. Moreover, the instance $Q_j a_{r+1}$ is non-negatively relevant to each individual κ^i $(i \neq j)$. (It need not be positively relevant. The new instance of Q_j changes the degree of disbelief of κ^i or $\{\kappa^k \mid k \neq i\}$ only if κ^j is the best hypothesis or κ^i is the best and κ^j the second best hypothesis according to $\rho_{\mathbf{r}}$.) Thus, our account of the confirmation of laws does justice to enumerative induction.

This is what I can offer as a translation of de Finetti's results into the framework of ranking functions. I find the translation basically plausible, and it strongly suggests following its course. One should characterize the class of ranking functions which represent mixtures of laws, and one should inquire the extent to which the representation is unique (for instance, there is an obvious one-one correspondence between the $\kappa_{\mathbf{r}}$ and the $\rho_{\mathbf{r}}$ in the above mixtures). One should look at de Finetti's representation results for the infinite as well as for the finite case (recall the finiteness assumption made in this paper). The ranking analogue to de Finetti's notion of partial exchangeability would be particularly interesting. And so forth.[24]

On the other hand, the translation still looks artificial and quite detached from actual practice. For instance, if min r_i is large, one would tend to say that all of the laws G^i are disconfirmed by $E(\mathbf{r})$ and to conclude that none of the laws holds. One might account for this point by defining some κ^0 representing the belief in lawlessness, by mixing it into κ_0, say with the weight $\rho_0(\kappa^0) = q$, and by finding then that as soon as min $r_i > q$ we have $\rho_{\mathbf{r}}(\kappa^i) = 0$ only for $i = 0$. This is definitely not the only way or place to improve upon the basic result.

So, there is certainly a lot of work to do in order to extend the proposal and to apply it to more realistic cases. But the message should be clear already from the basic case just explained. The theory of mixtures provides a clear account of what it means to confirm and disconfirm not only propositions, but also inductive attitudes such as ranking functions representing belief in laws. Hence, at least in this respect, my account does not leave a principal gap. Still, it is obvious

that this paper only begins to redeem the claim of its title: laws are persistent inductive schemes.

NOTES

1. This paper is similar in scope to Spohn (2002). That paper extends the analysis to *ceteris paribus* laws, whereas the present paper more strongly emphasizes the main point. Thus, my debts concerning the other paper extend also to this paper. Moreover, I am indebted to Eckehart Köhler for his very valuable comments and to Nils-Eric Sahlin for advice concerning Frank Ramsey's position.
2. However, the theory of induction takes a probabilistic turn in Lange (2000, ch. 4), a move with which I disagree, as I shall explain below.
3. For instance, Lange (2000) endorses these lessons when explaining what he calls the root commitment concerning the inductive strategies associated with laws.
4. For instance, Otte (1981) nicely documents how probabilistic theories of causation go astray with extreme 0-1-probabilities. Hence, the present attitude, as represented, e.g., by Pearl (2000), is to independently assume some account of deterministic causation, say, in terms of functional equations, which is then reflected in, but not defined by, 0-1-probabilities.
5. One may think that the incompleteness of Bayesianism is filled by Popper measures. However, if one combines the lessons of Spohn (1986) and (1988), it is clear that Popper measures are just as incomplete as is AGM belief revision theory. So, even the Bayesian has reason to buy into ranking theory as suggested in Spohn (1988), sect. 7.
6. $\overline{A}$ is the complement or the negation of A.
7. I apologize for the double negation; after a while one gets used to it.
8. Consistency and deductive closure have been often attacked and equally often defended. The issue of logical omniscience is indeed highly problematic. But we have already decided the issue by taking propositions as objects of doxastic attitude. I do not see viable alternatives.
9. It is obvious that one has various options at this point. For instance, in Spohn (1988) I still took the range to consist of arbitrary ordinal numbers, but the advantages of this generality did not make up for the complications. By contrast, Hild (t.a., sect. 3.2) does not extend the law of disjunction to the infinite case and is thus free to adopt non-negative reals as values.
   It is also obvious that the issue about infinite disjunctions is closely related to the discussion of the Limit Assumption in Lewis (1973, sect. 1.4). Without that assumption, it may happen that "if A were the case, then B_i would be the case" is true for infinitely many B_i which are jointly unsatisfiable. Lewis finds reason to accept this state of affairs. I prefer to accept the Limit Assumption instead.
10. I believe that if epistemologists talk of justification and warrant, they should basically refer to this relation of A being a reason for B; cf. Spohn 2001. That's, however, a remark about a different context.
11. A relevant argument is provided by the so-called problem of old evidence. The problem is that after having accepted the evidence, it can no longer be confirmatory. However, this is so only on the basis of the first condition. According to the second condition, learning about A can never change what is confirmed by A, and hence the problem does not arise. This point, or its probabilistic analogue, is made by Joyce (1999, sect. 6.4) with the help of Popper measures.
12. Note, by the way, that this would also hold for an infinite domain of quantification. Hence, for ranking theory there is no problem of null confirmation for universal generalizations which beset Carnap's inductive logic.
13. I am using here a technical notion of positive instance: a is a positive instance of G iff G_a, i.e. $Pa \rightarrow Qa$, is true. If $Pa \wedge Qa$, a positive instance in the intuitive sense, would be learnt, the beliefs would change, of course (at least given our assumptions that nothing except G is believed in κ).

14. "Resilient" might be an appropriate term as well, but I do not want to speculate whether this would be a use of "resilient" similar or different to the one introduced by Skyrms; cf., e.g., Skyrms(1980).
15. Here, $G_{\neq a}$ stands for $\wedge x(x \neq a \rightarrow G_x)$. Note that we have $\kappa(\neg G \mid \neg G_a) = 0$ according to both the persistent and the shaky attitude, simply because $\neg G_a$ logically implies $\neg G$.
16. However, it is obviously a difficult matter to describe the kind of belief change that has taken place in the change of attitude towards Bode's law. It does not appear to be just a kind of conditionalization. The ideas developed in section 5 might perhaps help here.
17. Cf., e.g., the overview in Lange 2000, pp. 11f.
18. Sahlin (1997, p. 73) explains that Ramsey's view of general propositions was obviously influenced by Hermann Weyl.
19. See Sahlin (1991, pp. 138ff.) for more details.
20. And he continues to explain the differences between the older 'inference licence' literature and his position which hide in the reference to "certain objective properties such as reliability".
21. But see my account of causal explanation in terms of ranking functions in Spohn (1991).
22. See, however, Spohn (1993).
23. I am still jumping between sentences and the corresponding propositions as seems convenient to me.
24. See, e.g., the rich results collected in the papers in Carnap, Jeffrey (1971) and Jeffrey (1980).

REFERENCES

Carnap, Rudolf / Richard C. Jeffrey (Eds.), *Studies in Inductive Logic and Probability, Vol. I*, Berkeley: University of California Press 1971.

Cartwright, Nancy, *Nature's Capacities and Their Measurement*, Oxford: Clarendon Press 1989.

de Finetti, Bruno, "La Prévision: Ses Lois Logiques, Ses Sources Subjectives", *Annales de l'Institut Henri Poincaré* 7, 1937. Engl. translation "Foresight: Its Logical Laws, Its Subjective Sources", in: Henry E. Kyburg jr. / Howard E. Smokler (Eds.), *Studies in Subjective Probability*, New York: John Wiley & Sons 1964, pp. 93-158.

Gärdenfors, Peter, "An Epistemic Approach to Conditionals", *American Philosophical Quarterly* 18, 1981, pp. 203-211.

Gärdenfors, Peter, *Knowledge in Flux*, Cambridge, Mass.: MIT Press 1988.

Halpern, Joseph Y., "Conditional Plausibility Measures and Bayesian Networks", *Journal of AI Research* 14, 2001, pp. 359-389.

Hild, Matthias, *Introduction to Induction. On the First Principles of Reasoning*, to appear.

Jeffrey, Richard C. (Ed.), *Studies in Inductive Logic and Probability, Vol. II*, Berkeley: University of California Press 1980.

Joyce, James M., *The Foundations of Causal Decision Theory*, Cambridge: Cambridge University Press 1999.

Köhler, Eckehart, "Physical Intuition as Inductive Support. Comments on Wolfgang Spohn's 'Laws are Persistent Inductive Schemes'", this volume 2003, pp. 151-167.

Lange, Marc, *Natural Laws in Scientific Practice*, Oxford: Oxford University Press 2000.

Lewis, David K., *Counterfactuals*, Oxford: Blackwell 1973.

Nayak, Abhaya C., "Iterated Belief Change Based on Epistemic Entrenchment", *Erkenntnis* 41, 1994, pp. 353-390.

Niiniluoto, Ilkka, "Inductive Systematization: Definition and a Critical Survey", *Synthese* 25, 1972, pp. 25-81.

Otte, Richard, "A Critique of Suppes' Theory of Probabilistic Causality", *Synthese* 48, 1981, pp. 167-189.

Pearl, Judea, *Causality. Models, Reasoning, and Inference*, Cambridge: Cambridge University Press 2000.

Popper, Karl R., *Logik der Forschung*, Wien: Springer 1934.

Ramsey, Frank P., "General Propositions and Causality", 1929, in: Frank P. Ramsey, *Foundations. Essays in Philosophy, Logic, Mathematics and Economics*, ed. by David H. Mellor, London: Routledge & Kegan Paul 1978, pp. 133-151.

Sahlin, Nils-Eric, "Obtained by a Reliable Process and Always Leading to Success", *Theoria* 57, 1991, pp. 132-149.

Sahlin, Nils-Eric, "'He Is No Good For My Work'. On the Philosophical Relations Between Ramsey and Wittgenstein", *Poznan Studies in the Philosophy of Science and the Humanities* 51, 1997, pp. 61-84.

Skyrms, Brian, *Causal Necessity*, New Haven: Yale University Press 1980.

Spohn, Wolfgang, *Eine Theorie der Kausalität*, unpublished Habilitationsschrift, München 1983.

Spohn, Wolfgang, "The Representation of Popper Measures", *Topoi* 5, 1986, pp. 69-74.

Spohn, Wolfgang, "Ordinal Conditional Functions. A Dynamic Theory of Epistemic States", in: William L. Harper / Brian Skyrms (Eds.), *Causation in Decision, Belief Change, and Statistics*, vol. II, Dordrecht: Kluwer 1988, pp. 105-134.

Spohn, Wolfgang, "A Reason for Explanation: Explanations Provide Stable Reasons", in: Wolfgang Spohn / Bas C. van Fraassen / Brian Skyrms (Eds.), *Existence and Explanation*, Dordrecht: Kluwer 1991, pp. 165-196.

Spohn, Wolfgang, "Causal Laws are Objectifications of Inductive Schemes", in: Jacques-Paul Dubucs (Ed.), *Philosophy of Probability*, Dordrecht: Kluwer 1993, pp. 223-252.

Spohn, Wolfgang, "Begründungen a priori – oder: ein frischer Blick auf Dispositionsprädikate", in Wolfgang Lenzen (Ed.), *Das weite Spektrum der Analytischen Philosophie. Festschrift für Franz von Kutschera*, Berlin: de Gruyter 1997, pp. 323-345.

Spohn, Wolfgang, "Wo stehen wir heute mit dem Problem der Induktion?", in: Rainer Enskat (Ed.), *Erfahrung und Urteilskraft*, Würzburg: Königshausen & Naumann 2000a, pp. 151-164.

Spohn, Wolfgang, "Deterministic Causation", in: Wolfgang Spohn / Marion Ledwig / Michael Esfeld (Eds.), *Current Issues in Causation*, Paderborn: Mentis 2000b, pp. 21-46.

Spohn, Wolfgang, "Vier Begründungsbegriffe", in: Thomas Grundmann (Ed.), *Erkenntnistheorie. Positionen zwischen Tradition und Gegenwart*, Paderborn: Mentis 2001, pp. 33-52.

Spohn, Wolfgang, "Laws, Ceteris Paribus Conditions, and the Dynamics of Belief", *Erkenntnis* 57, 2002, pp. 373-394.

Fachbereich Philosophie
Universität Konstanz
D-78457 Konstanz
Germany
wolfgang.spohn@uni-konstanz.de

ECKEHART KÖHLER*

PHYSICAL INTUITION AS INDUCTIVE SUPPORT

Comments on Wolfgang Spohn's "Laws are Persistent Inductive Schemes" †

I. GENERAL REMARKS

Wolfgang Spohn's paper I think clearly places him in the forefront of research on the theory of lawlikeness, and simultaneously on the methodology of inductive reasoning.

I find Spohn's entire approach quite sympathetic. I completely agree with Spohn's assessment that previous analyses of lawlikeness have still not achieved satisfactory results. For example, although both Spohn and I agree that counterfactuals reveal important factors of inductive reasoning, we share reservations about using them to try to characterize lawlikeness. Our deepest agreement is on what I shall call the "epistemologizing" of logic – or, to use more Carnapian terminology, the "pragmatization" of logic; Simon might call it "behavioralizing". For a long time, logicians, mathematicians (most of all those in the Hilbert tradition), and particularly analytic philosophers in the Fregean tradition have tended to suppress all reference to thought, perhaps for fear of being accused of psychologism, among other reasons. (I try to clarify this with an analysis of the subjective–objective distinction in §II below.) But I am, and apparently also Wolfgang Spohn is, convinced that it is crucial to consider inference as something actually representing (heuristic) rules applied by agents constrained by practical interests in real-life situations. What are called heuristics in artificial intelligence (AI) are just the "doxastic attitudes" that Spohn makes out in his §4; *cf.* my §V below. Therefore I find Spohn's approach to understanding lawlikeness within the context of doxastic dynamics absolutely sound.

II. SPOHN'S MAIN IDEA

Spohn's main characterization of lawlikeness is its doxastic "persistence" (§3). Strength of belief is measured by ranks, and on Spohn's model, ranks for (belief in) laws change in small jumps (in his finite models), depending on changes in evidence – and changes in "attitude". The distinction between "accidental" (or "empirical", in Kant's terminology) and "necessary" ("apodictic") is that be-

F. Stadler (ed.), Induction and Deduction in the Sciences, 151–167.
© 2004 *Kluwer Academic Publishers.*

tween "shaky" and "persistent" belief. Spohn holds an agnostic position on what ranking *interval* should count as measuring the distinction, perhaps in a behaviorist attitude, avoiding normativity when *describing* behavior. However, the fact that persistence is reasonably measurable at all makes Spohn's approach attractive.

Prima facie, a reclassification of an "accidental" law as a "necessary" one (§ IV below) ought to involve a big jump in ranking. If this is plausible, the question arises of *how big* the jump should be: what threshold or just-noticeable difference indicates that a law is held to be "necessary" and has achieved *sufficient* persistence? If the distinction is taken seriously enough, we might even consider an *infinite jump*, thus leading us to non-Archimedean rankings. In this connection, the question also arises of how much of Spohn's approach could be simulated by the Hintikka–Niiniluoto–Kuipers methods of (non-Archimedean) inductive probability distributions? (*Cf.* Kuipers 1978.) But Spohn warns us that his rankings are inherently superior to probabilistic methods. Popper would be pleased to hear *this* – but less pleased to hear that Spohn proposes an inductive method (§4).

A crucial step in Spohn's approach is to apply de Finetti's exchangeability theorem to characterize the confirmation of laws (§5). Since de Finetti's approach uses subjective probabilities, it of course needs adaptation to Spohn's doxastic rankings, but the approach still remains "subjective". Now the term "subjective" is sometimes perniciously ambiguous, and warnings are in order. Not that Spohn makes any errors here, since rationality theorists working with the known methods of subjective probability (and utility) are well acquainted with these concepts and avoid simple-minded errors. But naïve philosophers are still prone to confusion about subjectivity. In Köhler (2001, §4.3) and (2004, §9.1), I try to clear the situation up. One meaning of "subjective" is used to say something is *mental* (*i.e.* referring to a representational framework involving one of the two psychical faculties of perception or cognition, rather than merely *physical*); the other meaning is used to ascribe susceptibility to error to an observational or inferential claim, typically caused by highly variable factors assignable to individual agents (and therefore typically differing from person to person). It is important to emphasize that subjective probability theories, as well as philosophical and AI theories of doxastic dynamics or epistemic revision, are *normative* and hence make the claim to *perfection*. Even when they deal with error-detection and error-correction, the rules for this are themselves assumed to be error-free! Accordingly, if one meaning of "objective" is reliability and validity guaranteed by *freedom from error*, then de Finetti's and Spohn's theories are actually quite objective in this sense. If, on the other hand, we consider instead "subjective" theories of probability (logical inference, or mathematics, *etc.*) as theories about minds, we can quickly transform them into "objective" theories by simply rewriting subjective rules into objective propositions using operators in an "object theory". (We merely go from a metatheory which makes mental operations or states explicit to an object theory where they are only implicit.) For

example, every proposition stating the truth of another proposition A is "subjective"; but $T(A)$ can trivially be recoded into an "objective" (object-language) proposition, e.g. $A \equiv (0 = 0)$. So much for objective subjectivity.

Before I come to my main point, a brief remark on counterfactuals. Be skeptical about them, but don't discount them out of hand. First define confirmation and disconfirmation of propositions and laws doxastically, then interpret these metatheoretical attributes as object-theoretical doxastic modal operators. Then find out what modal logic they obey and see what happens. Although our epistemic intuition governing doxastic dynamics should or can be considered primary, *never* think they are *infallible*! It may be that intuitions about the modal-logic behavior of the doxastic modal operators can serve as *corrections* of doxastic attitudes at particular points. There is a good precedent for such a procedure: the interpretation of *logical proofs* as necessity operators by Löb and Boolos deepens our insight about Gödel's theorems, for example; see Boolos (1979), (1993). Although this is not uncontroversial, it is undeniably interesting, and can and already has redounded to profit proof theory.

III. TWO EXAMPLES CONCERNING EINSTEIN

I now come to my main point: I want to consider how accidental laws can effectively be distinguished from necessary laws. I conjecture that normal observational data only rarely influence this classification; but *background heuristics* do, often *prior to* any data collection. This may at first sound to be a hopeless task, considering that heuristics are so manifold and difficult to grasp. (I realize full well that research in AI is only at a bare beginning on heuristics.) Nevertheless, I feel certain that they play the main role in deciding just such issues as persistence of belief, or whether to classify a law as "necessary" rather than "accidental".

To consider an example concerning Special Relativity Theory, recall that Einstein (in retrospect very surprisingly) may *not* have known about the Michelson–Morley experiment of 1887 when writing his 1905 paper on Special Relativity, whose central idea was based on a heuristic (a famous "gedanken-experiment") about an observer's trying to catch up with a ray of light. Holton (1969) tries to explain why Einstein (1905) neither mentioned nor used knowledge of the Michelson–Morley experiment. According to "standard" methodology of science, as soon as the experiment was published all theoretical physicists should have *immediately* abandoned the classical ether theory requiring a specific medium for transmitting electromagnetic forces. But in fact, most theoretical physicists "persisted" in the perhaps shaken but apparently irreplaceable ether theory – they were wedded to a certain heuristic (Spohn's "doxastic attitude"). It took a revolutionary like Einstein to *break out* of the old heuristic and propose a better one. Pais (1982, Ch. 6a, b) expertly shows, *pace* Holton, that Einstein knew about the Michelson–Morley experiment after all – through

Lorentz; but he nevertheless agrees that Einstein didn't "need" it, *already* having been motivated by a prior, quite persistent heuristic rejection of absolute motion; Pais then adds some caveats against what Spohn might call "rash persistence" (1982, Ch. 8, *p.* 172):

> In a letter to a historian [F.C. Davenport], written a year before his death [9 Feb. 1954], Einstein expressed himself for the last time on this subject: 'In my own development, Michelson's result has not had a considerable influence. I do not remember if I knew of it at all when I wrote my first paper on the subject (1905). [Pais shows that Einstein *did* know.] The explanation is that I was, for general reasons, firmly convinced that there does not exist absolute motion and my problem was only how this could be reconciled with our knowledge of electrodynamics. One can therefore understand why in my personal struggle Michelson's experiment played no role, or at least no decisive role'.
>
> Why this need not to remember or, at best, to underplay this influence?
>
> Just over twenty years before Einstein wrote this late letter, just under twenty years after his creation of the special theory, he gave a lecture at Oxford entitled 'On the Method of Theoretical Physics' (Einstein 1933), in the course of which he said, 'It is my conviction that pure mathematical construction enables us to discover the concepts and laws connecting them, which give us the key to the understanding of the phenomena of Nature.' It seems to me that here Einstein grossly overestimates the capabilities of the human mind, even of one as great as his own. It is true that the theoretical physicist who has no sense of mathematical elegance, beauty, and simplicity is lost in some essential way. At the same time it is dangerous and can be fatal to rely exclusively on formal arguments. It is a danger from which Einstein himself did not escape in his later years.

Pais goes on to speculate on the powerful effect of the "ideal axiomatic structure" of the kinematics of Special Relativity when Einstein suddenly discovered it after a discussion with Besso; and on the effect Riemannian geometry had, taught to Einstein by Minkowski in Zurich but unused before developing General Relativity, when it was recalled by Großmann to help Einstein. But Einstein and Pais need to be corrected on one point: neither of these were "pure mathematical constructions"!! They were drenched with physical meaning – for geometry is physics, not pure mathematics![1] Physical heuristics are to be distinguished from purely mathematical heuristics by the feature that the former must in some way be related to observable physical states – or at least theoretically possible measurements in gedankenexperiments. Einstein's remarkable formulation in (1933) contradicts his own life-long insistence that physical theory-construction be "physical" and not purely mathematical. (*E.g.*, Einstein doubted the usefulness of Gödel's solutions allowing time-travel as being "unphysical" – meaning that it couldn't be physically realized.)

Even more interesting is the famous test of General Relativity authorized by the British Astronomer Royal, Sir Frank Dyson, and carried out by Eddington in 1919, which was perhaps less subject to reinterpretation than was the Michelson–Morley experiment. This observation of the deflection of the light of a star near the corona of the sun during a total eclipse made Einstein a world-famous man and clinched his reputation as an equal of Newton – particularly in Great

Britain; *cf.* Pais (1982, Ch. 16). A student asked Einstein at the time what he would have done if the observation had turned out *negatively*, and Einstein replied with firm conviction that he would have continued as before – he would have persisted in his conviction. Clark (1971, *p.* 230):

> Looking back, both Einstein and his colleagues were apt to harp on his inner certainty. Thus Ilse Rosenthal-Schneider, one of his students, remembers how, as the two of them were discussing a book which raised objections to his theory, Einstein reached for a telegram lying on the windowsill and handed it to her with the words: "Here, this will perhaps interest you." "It was," she has written, "Eddington's cable with the results of the measurement of the eclipse expedition. When I was giving expression to my joy that the results coincided with his calculations, he said, quite unmoved, 'But I knew that the theory is correct,' and when I asked, what if there had been no confirmation of his prediction, he countered: 'Then I would have been sorry for the dear Lord – the theory *is* correct.'"

Incidentally, this seems to imply that Einstein would not even have considered a negative result as a disconfirmation (to say nothing of "falsification") of his theory, having given very strong persistence (or consistently high ranking) to his heuristic principles over and above isolated observations. Eddington's data are quite clear and difficult to re-interpret Also to be kept in mind, however, is that the interpretation of observations which are *heavily* theory-laden becomes additionally shaky, and hence their influence on doxastic dynamics becomes more subtle.[2]

IV. A CRUCIAL EXAMPLE: BODE'S "LAW"

A most famous test case for Spohn, which is more classical but no doubt trickier to interpret than Einstein's gedankenexperiments, is that of Titius's old empirically discovered law, confirmed and made famous by Bode, of the logarithmic ("geometric") distribution of planets in the solar system. In the eighteenth century, when considering distances of the planets from the sun, and influenced by the persisting interest in Napier's logarithmic functions, which were wonderfully helpful in speeding multiplication, Titius discovered the interesting fact that each planet is almost exactly twice as far from the sun as the next closer planet.[3] The authoritative source on the history and theory of Bode's Law is Nieto (1972); the easily accessible Murray (2002) relates some of the latest developments. No explanation was known for this law at the time, and even authoritative current references like the *Encyclopedia Britannica* deny any explanation for it; it had "come out of the blue". There was much debate about whether the nice distribution should even be dignified by calling it a law. It was unknown whether it was a regularity, as no other solar systems were known.[4] Clearly, Bode's "Law" is a great candidate for an "accidental law": attitudes toward it had always been shaky, it was never clear how strongly it deserved assent, strong calls came for finding an underlying dynamic explanation for it.

This attitude remained unchanged for several centuries, because of the sketchy (or should I say, in deference to Spohn) "shaky" acquaintance with celestial dynamics. Newton of course was not able to solve the famous *n*-body problem, whose mastery was required to prove that the solar system was stable. Laplace later attempted this proof, but it is now accepted that he unjustifiably rounded off just the factors responsible for unstable perturbations of planetary orbits. (However, he did explicitly allow that a comets with unstable orbits are quite possible, and that they couls easily cause catastrophes.) Based on advances in celestial dynamics going back to Poincaré and Newcomb in the nineteenth century, two researchers in the twentieth century, first Brown (1932) and then Hills (1970), suggested *explaining* Bode's Law as representing a state of "dynamic relaxation", meaning that the solar system is at its most stable point when planets are distributed logarithmically. Ovenden (1976) proposed a "principle of least interaction action" (*cf.* also Roy & Ovenden 1954). Bass (1974, referring to a paper of 1958) proposed a quite similar principle of "least mean absolute potential energy".

Then Kolmogorov (1954) made the momentous discovery that catastrophes in planetary systems are all too likely when he found that the minor terms in Newcomb's series used to solve the *n*-body problem for planetary motion in Newtonian mechanics were in fact easily subject to random, unpredictable perturbations which could quickly lead to turbulence; *cf.* Arnol'd (1997). (In the meantime, the buzzword "chaos" is applied to such turbulance – also called "catastrophe", to the dismay of Shapley and many other astronomers sharing Newton's dream of an eternally stable solar system.) Kolmogorov was uniquely qualified to make his discovery because of his mastery of several specialties simultaneously: probability theory *and* deterministic mechanics *and* topology. Subsequently, in the later fifties, Siegel & Moser (1971) and Moser (1973) further extended mathematical knowledge of the divergent Newcomb series.

That "deterministic" systems like those obeying Newtonian mechanics should be subject to unpredictable catastrophes came as a real shock. But Poincaré had already discovered chaos in 1890, and a prominent member of the Vienna Circle, Philipp Frank (1934), who knew quite a lot about continuum mechanics and its mathematics, spelled out the philosophical consequences in scintillating detail reminiscent of the iconoclastic Mach: even outside of Quantum Physics the notion of determinism (a universal law of causality) has little practical meaning and is virtually a tautology.

Next came pioneering computer simulations of planetary-system evolution by Hills (1969). Wisdom (1981), finding even more advanced numerical methods and using even faster computers, achieved the breakthrough of repeatedly observing chaos (or catastrophes) in computer-simulated planetary orbits, splendidly confirming Kolmogorov and putting the dream of Newton in an absolutely stable and orderly cosmos to rest in a way most astronomers could not have previously imagined. All experts now agree that even apparently stable orbits will

occasionally experience unpredictable, chaotic disturbances: Roy (1991), Scheck (1999), Laskar (1994), Sussman & Wisdom (2001).

Now the irony: the very same numerical simulations that confirmed catastrophism, and even the earliest simulations by Hills, displayed great orderliness: evolving planetary systems always tend strongly to satisfy Bode's Law: no matter what initial conditions were presupposed, planets settle into orbits distributed according to the law! Most recently, Laskar (2000) provides a dynamical explanation for this remarkable result; and Hayes & Scott (1998) discuss statistical features relating randomly selected planetary systems with Bode's Law. All of this new work confirms Bass and Ovenden: planets will typically remain in stable orbits distributed according to Bode's Law for long periods, but they are subject to occasional hard-to-predict instabilities, which in turn quickly "relax" back into stable orbits satisfying Bode's Law again – although the order of planets perhaps may have been flipped. It has even been worked out how Venus could have switched step-wise from an orbit between Jupiter and Saturn to its present (quite circular!) orbit inside the Earth's.

These results *seriously embarrass* certain highly prominent astronomers, ignorant of the progress in celestial dynamics, *viz.* Harlow Shapley and Carl Sagan, who had wrongly persisted in believing in Newton's dream of a billion-year-long stability of the solar system (which Newton desired for religious and possibly political reasons). Caught in their mistaken ideology, they had violently rejected the claim of Velikovsky (1950) that Venus is a new planet having had an unstable, cometary trajectory in historical times before settling down in its present orbit.[5] Bass (1974) pointed out that advances since Kolmogorov (1954) and Hills (1970) strip away the main argument of Velikovsky's opponents.

Now we can no longer consider Bode's Law "accidental" at all, because we now know it represents an energetic equilibrium point inherent in all planetary systems – without implying absolute stability, however! It is *not absolute*, because it tolerates occasional chaotic disturbances when it is temporarily "violated" (note that Spohn emphasizes in his §4 Discussion that violated laws need not thereby suffer in their ranking). But it *is universal* in the sense of holding for all possible planetary systems, and all those observed up to now adhere to it; which obviously beefs up its ranking. In any case, with reference to Spohn's doxastic dynamics, we may say that those astronomers exposed to the most recent theory were *now* suddenly confronted with surprising data from a newly discovered planetary system showing a *non*-logarithmic distribution, they would nevertheless persist in believing in Bode's Law; they would decide that the planetary system observed is merely undergoing a short-term catastrophe of the sort Kolmogorov showed to be all too likely. Spohn's doxastic dynamics has no problem dealing with this discovery of previously unknown logical connections between laws: Newton's Laws would already have high values for persistence of belief, and adding the mathematical discoveries of Kolmogorov, Wisdom and Laskar would then transfer those high values to all their logical consequences – including Bode's Law.

On the other hand, it seems impossible for Spohn's ranking functions to distinguish between accidental and necessary laws purely on the basis of observational data and absent any information contained in heuristics or in accepted laws. Spohn concedes this by his otherwise unmotivated distinction (§4) between "belief in laws" and "belief in propositions", where he first suggests introducing "doxastic attitudes" to bridge his embarrassment, and further suggesting "metaphorically" that confirmation of laws involves "second-order inductive attitudes". Since my heuristics are quasi-logical norms, like Carnap's meaning postulates, which guide theory-valuation, this seems to explicate their "second-order" status, giving them just the independence from bare data ("brute facts"?) that Spohn claims for them. On the other hand, I see no reason for *not* considering laws to be propositions. The Vienna Circle already tried that, only to give it up again. Neurath would immediately point out that many protocol sentences can be better entrenched than some laws. As Neurath insisted, with reference to Duhem, laws are on the same level as protocol sentences: either can confirm or disconfirm the other. What's special about laws is our doxastic attitude toward them reflecting our heuristics, and *these* are quasi-independent.

In the absence of explicit knowledge about heuristics, we might instead try "training" the ranking function to track the efficiency of laws, reflected *i.e.* in their explanatory power (*e.g.* in the sense of Niiniluoto & Tuomela 1973). However, such "training" would certainly merely reproduce just the heuristics assumed to be absent. Presumably, most "normal" science (in the sense of Kuhn) relies on not-very-dynamic rankings of laws accepted by the "guild" (as with Shapley and Sagan); but meta-theoretical heuristics clearly guide abduction when developing new theories (hence induction).

V. PERSISTENCE REFLECTS HEURISTICS

Spohn provides crucial help in clarifying the inductive behavior of ranking functions by adapting de Finetti's exchangeability theorem (intended for subjective probabilities) to the structurally similar ranking functions (§5). In previous articles, Spohn (1991) and (1993) provides details on the general "behavior" of explanation and induction of causality claims.

What Spohn's approach does not (yet) include are norms actually assigning (at least relative) ranks to laws and observations. This he must do to show that accepted laws are (or *should be*) significantly more persistent than "accidental" generalizations or (sets of) observations. What we need is ranks for heuristics of theory discovery and evaluation. Actually, these are not entirely novel in inductive reasoning, as something close is already well known and widely accepted in inductive logic and statistics – disguised as constraints on sample spaces! Carnap (1952) has explicated these in semantic guise as "meaning postulates", displaying observational or measurement terms as governed by (para-)logical relations which are typically assigned a probability equal to 1 (*i.e.* treated as "practically

certain"); examples include the famous phenomenological "facts" (beloved of Husserl and Wittgenstein) that nothing green is red, *etc.*; and that nobody is taller than himself. All statistical investigations impose similar constraints on events and outcomes which are not logical in the sense of being independent of the empirical meaning of terms, since they're derived from the nature of the observational setups used, *i.e.* of the classification schemes and relations used to describe data and "known" even before the data are collected.[6]

The heuristics I have in mind perhaps belong to an epistemic category somehow higher than Carnap's meaning postulates. They represent newly invented constraints guiding theory construction like those used by Einstein in developing Relativity; but also those reflecting "received opinion" in routine, "normal" science. They seem to be what Spohn (§4) glorifies as "doxastic attitudes"! They are individually identifiable, researchers get attached to them – they gain "fetish"-value, *i.e.* they represent a person's identity, they represent scientific *culture*. They are no doubt much less tractable than lower-level statistical constraints, and may not deserve a probability equal to 1 (Spohn's rankings avoid this anyway). But assuming Einstein was behaving rationally, our models of doxastic dynamics have to reflect the heuristics which motivated him to believe persistently in *both* his Special and his General Relativity Theories. Inductive logics assigning rankings therefore have to use lists of heuristics.[7] Such heuristic principles need to be *added* to standard inductive inference rules, as they have the effect of drastically narrowing down classes of theories considered by the theoretician (Simon's heuristics of *search*). Einstein had good reasons for using the heuristics he did, and if we analyze his reasons, we (or at least Pais) can usually justify or criticize each, formulating everything as meta-theoretical rules. Heuristic principles are rules with logical structure related to similar as well as to competing rules – to which axiomatic analyses can be applied *pro* and *con*, in ways typical of axiomatic science. Bolzano (1837, Vierter Theil) is the great pioneer at *describing* (or perhaps, like Hume, *prescribing*) such analyses, and his greatest successors are Polya (1954) and Simon (1977). Such principles will serve as a core of inductive reasoning. Discovering and refining heuristics will no doubt be hard work, but work in AI indicates it can be done. Newell & Simon (1956) were the first to embed heuristics in computer programs to accomplish logical inferences. Lenat (1982), Simon's student, has been very successful in developing heuristics for doing arithmetic on computers.

In my work on Gödel's philosophy of mathematics (Köhler 2001, 2002), I claim that heuristics such as Lenat's are in fact what mathematicians traditionally call intuition. My explication of mathematical intuition makes it a procedure which evaluates mathematical axioms or rules guiding mathematical proofs and calculations. As Gödel insisted, every concept or proposition, concrete or abstract, has intuitions associated with it. In physical work, we need to speak of "physical intuition" – *e.g.* Einstein's intuition of space-time; in ethics, of "moral intuition", *etc.* Intuition is thus primarily local – Einstein's gedankenexperiments on light beams and elevator rides are extremely "situational". This is why it is

not feasible to develop a *general* inductive methodology powerful enough to develop theories in all special areas (unless the general method already builds in all local knowledge). This becomes even more obvious when taking the *fallibility* of intuition into account as well: error detection and correction are most especially local. This feature of "locality" runs counter to the property of universality or generality which logicians hold so dear; but this should not surprise us if we consider the extreme "situationality" of the logic of sample spaces in statistics.

In any case, I am convinced that heuristics of theory choice and theory valuation are essential components of doxastic dynamics – and that they are logical and not empirical. Meaning postulates for statistical terms are logical, because they belong to the *norms* of executing observations and statistical inferences. Heuristics of theory development are similarly logical, because they belong to the *standards* scientists use for theory choice, thus helping to define knowledge.

VI. THE UNCRITICAL IRRATIONALISM OF POPPER'S ANTI-INDUCTIVISM

Finally I am ready to tackle Popper's skepticism about inductive reasoning. My motive is to buttress the rather "behavioral" approach of Spohn's ranking functions.

Popper knew, just as Hume knew, that everyone uses inductions fairly often. They are habits of thought which Hume even emphasized were strongly held. The usual interpretation of Hume is that mere habits of induction cannot warrant valid inference of conclusions from premises non-deductively. Hume, much more so than Popper, was deeply interested in cognitive psychology – indeed, he wanted to become the Newton of the mind; Hume was especially interested in the psychology of these strong habits of induction which humans use with obvious profit. To those who know Hume as the great skeptic of induction, it therefore comes as a surprise that he formulated rules which later became highly respected as pillars of inductive methodology. (Here we may ask a crucial question about Hume's methodology: was he proceeding *descriptively* or *prescriptively*? My guess is prescriptively – based on insight that they "must" be right! He was inventing or discovering them, not taking surveys of average behaviors.) These were the rules taken over by John Stuart Mill, which became known as Mill's Methods. In my view, Wilson (1997) provides convincing arguments for this interpretation of Hume. (Still, it would probably be closer to the truth to admit that Hume was probably blatantly inconsistent on the question of the rationality of induction, whatever his "main inclination" was; even more so to admit that Hume's writings led many readers to believe that Hume held induction to be irrational and even, most notably Kant, to reject induction. Could a great philosopher like Russell go completely astray in claiming Hume "destroyed empiricism" – the consequence of rejecting induction?)

At least from an empirical, psychological point of view, the feature of *strong plausibility* these rules enjoy is something they share with deductive methods as well. Let us consider the feeling of plausibility for a working mathematician. For him, it is in fact the very intuition he relies on for discovering conjectures and proofs, and for finding nice definitions of murky notions (for example, the famous definitions of dimension and curve by Menger 1923, 1925). It constitutes the very essence of mathematical knowledge, or at least skill. Both "objectivist" (Platonist) as well as "subjectivist" (intuitionist) schools in mathematics regard mathematical intuition as valid in assessing our mathematical work. In short, the feeling of plausibility (intuition) represents a type of observation, a "sixth sense", as Gödel called it, which supports and justifies our belief in the validity of mathematical rules and axioms. And just as observing physical objects with our senses is taken to justify objective knowledge of them – assuming the observations were made by trained people under "normal" circumstances –, so the subjective feelings of plausibility guiding mathematical research justifies claims of objective knowledge of theorems and calculations – again, assuming the mathematician is trained and working under "normal" conditions.

By now, my Popperian readers may start squirming, because it may be dawning on them how my argument will now proceed. It is that the *feeling of plausibility* that scientists feel when using *inductive* methods guiding their work is *just as* valid as intuition is for the mathematician when using *deductive* methods. Statistical intuition justifies acceptance of propositions (claims to objective knowledge) just as well as deductive intuition does. Above all, there is no reason for denying feelings of statistical plausibility the *power of validating* those rules of induction, of calling them reasonable or rational. This is widely done in all the many fields where normative theories or methods are developed, from decision theory to econometrics to accounting to law and to ethics. Exactly this view is promulgated by Wilson (1997) for Hume. How can Popper not agree? Popper continued to attack all inductive logics and methods of statistical inference to his death. He claimed to have "solved Hume's problem" of induction – but in fact he denied the solubility of the problem *tout court* when openly stressing that "I do not make predictions".

Early last century, Fisher (1925) and Neyman & Pearson (1933) developed the first medium-sophisticated methods of statistical hypothesis-testing, but these came just *after* Popper had already fixed his persistent conviction[8] that inductive method plays no role in scientific reasoning. Later on, by the seventies, the Bayesian approach of Ramsey, de Finetti and Savage caught on, and Popper violently fought that, too. Carnap's inductive logic he fought most of all.

In 1965, Rudolf Carnap gave a paper at the conference on philosophy of science in London organized by Lakatos; it was published as Carnap (1968). In his paper, he emphasized again, as he had already in the "Replies" in his Schilpp volume in 1963, that inductive logic required inductive intuition as its foundation and justification. He was led to this view – so Shimony told me – by considering how de Finetti's exchangeability theorem represents our epistemic intuitions in

the probability distributions we decide to accept. (Just the same tool Spohn has discovered to reveal his "doxastic attitudes"!!) Carnap was of course worried that his dear old empiricist friends, like Hempel, perhaps Jeffrey, and many others, would be rather upset at his stressing intuition as he did. And for good reason, because from an empiricist standpoint it is forbidden to use intuition to justify anything – empiricism had ever opposed the hubris of rationalist "insights", *e.g.* those of philosophers such as Descartes and Leibniz).

Carnap (1968, *p.* 259) first establishes the goal of inductive logic as prescriptive rather than descriptive:

> But I would like to distinguish two kinds of questions of degree of belief [equivalent to inductive probability], or two theories about degrees of belief. The one is a factual, empirical theory which belongs to psychology. It investigates under what conditions people have beliefs of such and such a degree. ... But in inductive logic, we are not concerned with the actual degrees of belief that people have and the causal connections between these degrees and other factors, but rather with the *rational* degree of belief. Thus inductive logic has the task of telling us how to arrive at values for our degrees of belief which we can defend as rational. ...

Carnap (1968, *p.* 265*f.*) then argues that intuition justifies inductive inference just as well as it does for deductive inference:

> ... In order to learn inductive reasoning, we must have what I call the ability of *inductive intuition*. My friends warned me against the use of the word 'intuition'. So I hasten to add that I do not mean by 'intuition', here 'inductive intuition', a source of knowledge that is infallible. I don't think there is anything infallible in human knowledge or activity. Certainly inductive intuition may on occasion lead us astray; the same holds for all other kinds of intuition, for example spatial intuition, which is used in geometry. Maybe you have the feeling that this mysterious inductive intuition is a rather dubious basis for such a serious enterprise as the construction of a system of inductive logic. If so, I would like to call your attention to the fact that the situation in deductive logic is exactly the same. If you had before you a person who was deductively blind, that is to say, unable to distinguish between valid and invalid deductive inferences, or between deductively valid and invalid statements, even in the simplest cases, then you could not do anything with him. You could not convince him of the validity of the *modus ponens* inference. You tell him, 'You just told me that it is raining now.' 'Yes,' he says, 'You told me before that if it is raining the lawn is wet.' 'Yes. So then you believe now that the lawn is wet.' 'Why should I?' he replies. ... But the essential point is this: you have to appeal to his deductive intuition in order to teach him deduction. Thus we see that the epistemological situation in inductive logic, as I described it earlier, is not worse than that in deductive logic, but quite analogous to it.

But Popper (1968, *p.* 296) responded by rejecting intuitions as a source of knowledge entirely, calling them (mere) "beliefs" in contrast to "criticizable statements-in-themselves in the sense of Bolzano" – ignoring that Bolzano himself, like Frege, wrote that knowledge of logic depends on "Einsicht" (insight) – the same as intuition! Still and all, Popper agrees that deductive inferences can

indeed be intuitive (even if that's "irrelevant" for science, which spurns mere beliefs), so it seems excessive for him to deny at the same time that *inductive inferences* are *ever* intuitive.[9] (As if the huge industry of statisticians and econometricians were totally unintuitive and entirely random!) Popper could not and his followers[10] cannot deny that Neyman–Pearson as well as Bayesian methods are *de facto* widely used and appear at least a little intuitive to the users. And animals clearly *use* induction as well, *e.g.* when learning by example. Moreover, thousands of economists, decision theorists and Operations Research specialists openly call Bayesianism their standard concept of rationality, the core methodology of the fabled *homo oeconomicus*. Altogether, if we count them all, there may be hundreds of thousands of trained users of statistics and econometrics and accounting who would be astonished if anyone were to say their methods were irrational. But Popper's position has exactly this consequence. (His position equally forced him to reject all forms of probability estimation, so he was blocked from determining any values for his beloved propensities.)

I think Carnap was right: the real solution to the "riddle" of induction is to attain a conviction in the *reasonableness* of the the rules of induction – just as we have faith in deduction because we are convinced by deduction rules. It's as simple as that.

Recall that Friedrich Nietzsche was famous for denying that morality could be justified by any arguments; for Nietzsche, no norms are valid: there is no question of "truth" or "falsity", no "right" or "wrong" when obeying norms of justice. Nietzsche preached "transvaluing" all values, he preached "living dangerously", breaking rules. He called himself a nihilist, and his philosophy was widely called irrationalist. Anyone like Popper who says statistical methods of inductive reasoning patiently developed over many decades by reasonable and skilled experts are *in principle* illegitimate and *nothing like them* could ever justify any belief is insofar irrationalistic, is nihilistic. Statisticians are desperately depended on by insurers, medical doctors, investment consultants, businessmen, *etc.*, for their skillful advice. To say their advice cannot be valid on principle is absurd.

In conclusion, if we agree intuition can justify our use of induction, this buttresses Spohn's approach of behaviorally measuring ranks of belief, since we can suppose that "persistence" (or "entrenchment") of beliefs and methods indicates their rational validity. In any case, I am sure that Wolfgang Spohn is a reasonable person, and I'm convinced he carefully consults his intuitions in developing his new and interesting doxastic methods. I hope he will consult the local intuitions of scientists in their various fields as well.

NOTES

* This paper was presented at the conference "Induction and Deduction in the Sciences" organized under the auspices of the European Science Foundation by the Institute Vienna Circle at the University of Vienna, 7–9 July, 2002. This conference took place immediately after a conference on Karl Popper in Vienna; I thank the organizers for giving me the opportunity to criticize Sir Karl without delay (§VI below) for having persisted in his well-known opposition to inductive reasoning.

† A paper similar in content, Spohn (2002), has appeared in *Erkenntnis* **57**.

1. Gauß, Helmholtz and von Neumann were absolutely convinced of this, contrary to Kant. Especially in Newton we may see that mechanics, as in his great predecessor, Archimedes, is a straightforward extension of Euclidean geometry, adding kinematics and a force law. Conversely, as Bolzano already noticed, Kant mistakenly took arithmetic to be *physical*, not realizing that counting does not involve time essentially.
2. Bayesians like Jeffrey (1965) try to take such problems concerning the fallibility of observation into account in their updating schemes by never assigning observation propositions probabilities equal to the customary 1, as Carnap used to do. Spohn's ranking functions avoid this problem *a priori*.
3. Contributing to the great interest in the discovery were the aesthetic and mystical connotations of the fact that the proportions between orbits of planets were that of the "Golden Mean", famous *e.g.* in Greek temples, in Renaissance architecture (also in Hindu and Buddhist tradition), which apparently confirmed Copernicus's and Kepler's Pythagorean fixation on the "harmony of the spheres" in the solar system. Moreover, the numerical sequences Titius generated represent Fibonacci sequences, giving additional impetus to admirers.
   Major exceptions are Neptune and the very eccentric Pluto (smaller than Mercury, hence not taken so "seriously"), and the unoccupied gap between Mars and Jupiter where the asteroid belt was later discovered. The latter was speculated by some, notably Ovenden (1973), to have once been occupied by a large planet named Phaethon (by Russians) or Aztex (by Ovenden) that disintegrated before historical times in a catastrophe; such speculation has not been approved of by most astronomers, because it can only be supported by tricky historical research – which, as the Velikovsky affair has shown, is extremely controversial.
4. However, it was soon enough confirmed for the moon systems of Jupiter, Saturn and Uranus, which Galileo had already recognized as mini-solar systems; and 20th century astronomy has brilliantly confirmed it as a regularity in the recently discovered planetary systems around distant stars, as well as in galaxies. (The moon systems all seem to be prone to chaos, however!)
5. Even before Velikovsky (1950) was published by Macmillan, Harlow Shapley of Harvard (father of Loyd Shapley, the game theoretician), the most influential American astronomer, began organizing an infamous campaign (*cf.* Juergens 1967) to censor and denounce Velikovsky, coercing Macmillan to drop his best-selling book. Shapley ignored earlier warnings by a leading expert on celestial dynamics, Brown (1932), against taking Laplace's "proof" too seriously. Sagan (1974), a student of Shapley's who had meanwhile become world-famous at Cornell, continued the scandal with erroneous "refutations" of Velikovsky; *cf.* Ginenthal (1994).
   Velikovsky's claim was based on extensive documentary evidence, primarily Middle Eastern, but also Chinese, supplemented by speculations about possible electromagnetic influences (further upsetting the experts, who didn't know about the Van Allen belt yet); and on worldwide ethnonological evidence – none of which had sufficient dignity to shake Shapley, of course. Only the Babylonian tables of sightings of Venus might have moved him – or at least those like Newton who value ancient documents –, but the British Museum refused to cooperate.
6. Of course, the ivory-tower philosophers Husserl, Wittgenstein and their followers made a serious mistake in trying to discover apodictic secrets about such *a priori* knowledge independently from familiarity with observational technology and hardware (including psychophysics), *and* independently from specific (conceptualized, axiomatic) theoretical contexts, *and* disregarding its fallibility! Instead, they deluded themselves into thinking unaided introspection could dis-

cover such deep secrets – as if introspection knew very much about physiology of organs, about instrument engineering, or about axiomatic details of theories.

7. The major guru of heuristics in methodology of science and in AI, who received a Nobel Prize for his work, is Carnap's student (and Lenat's teacher), Herbert Simon. See Simon (1977), Simon (1979), Newell, Shaw & Simon (1962), and Langley, Simon, Bradshaw & Zytkow (1987).
8. Under the influence of Jakob Fries, Leonard Nelson and Julius Kraft; *cf.* Hacohen (2000, *p.* 117*ff.*).
9. Popper (1968, *p.* 297): "My point is that there is no similar rule of inductive logic which is either intuitively valid or which stands up to criticism in the light of the history of science." Had Popper not heard of the mutual criticism between Fischer and Neyman, whose statistical test theories also constitute inductive logics? The community of statisticians, based on their intuitions, tended to accept Neyman's criticisms.
10. Watkins (1984, Ch. 9) was Popper's most loyal follower – until he realized that Lakatos and Salmon were right in their criticisms of Popper's anti-inductivism. Watkins appeals to the *utility* of (the success of) prediction rules; but the psychometrics of utility reveal that utility functions measure (in this case scientific) feelings of satisfaction (*e.g.* of "correctness"), so we are right back to plausibility and intuition of trained scientists. PS: Clearly intuition also has to respect pragmatic aspects like likelihood of success.

REFERENCES

Arnol'd, Vladimir I. (1997): *Mathematical Aspects of Classical and Celestial Mechanics* (with contributions by A.I. Neishtadt), Springer-Verlag, Berlin.

Bass, Robert (1974): "'Proofs' of the Stability of the Solar System", *Pensée*, No. 8.

Bolzano, Bernard (1837): *Wissenschaftslehre*, vol. **III**, Sulzbach, Bavaria; edited by Alois Höfler (*post.*) and Wolfgang Schultz and reprinted by the Verlag von Felix Meiner, Leipzig 1930.

Boolos, George (1979): *The Unprovability of Consistency*, Cambridge University Press, Cambridge; rewritten and extended in (1993).

Boolos, George (1993): *The Logic of Provability*, Cambridge University Press, Cambridge.

Brown, Ernest W. (1932): "Observation and Gravitational Theory in the Solar System", *Publications of the Astronomical Society of the Pacific*, 44, 21-40.

Carnap, Rudolf (1952): "Meaning Postulates", *Philosophical Studies*, **3**, 65-73.

Carnap, Rudolf (1968): "Inductive Logic and Inductive Intuition", in Imre Lakatos (ed.) (1968): *The Problem of Inductive Logic. Proceedings of the International Colloquium in the Philosophy of Science, London, 1965*, Vol. **2**, North-Holland Publ. Co., Amsterdam.

Clark, Ronald W. (1971): *Einstein – The Life and Times*, World Publishing Co., New York.

Earman, John, and Glymour, Clark (1980): "Relativity and Eclipses: The British Eclipse Expeditions of 1919 and Their Predecessors", *Historical Studies in the Physical Sciences*, **11**, 49.

Einstein, Albert (1905): "Zur Elektrodynamik bewegter Körper", *Annalen der Physik*, ser. 4, **17**, 891-921.

Einstein, Albert (1915): "Erklärung der Perihelbewegung des Merkur aus der allgemeinen Relativitätstheorie", *Preußische Akademie der Wissenschaften*, Pt. 2, 831-839.

Einstein, Albert (1916): "Die Grundlage der allgemeinen Relativitätstheorie", *Annalen der Physik*, ser. 4, **49**, 769-822.

Einstein, Albert (1917): "Kosmologische Betrachtungen zur allgemeinen Relativitätstheorie", *Preußische Akademie der Wissenschaften*, Pt. 1, 142-152.

Einstein, Albert (1933): *On the Method of Theoretical Physics*, Oxford University Press, Oxford.

Fisher, Ronald, A. (1925): *Statistical Methods for Research Workers*, Oliver & Boyd, London.

Frank, Philipp (1934): *Das Kausalgesetz und seine Grenzen* [Schriften zur wissenschaftlichen Weltauffassung **6**], Springer-Verlag, Vienna; reprinted under the editorship of Anne Kox (1988), Suhrkamp Taschenbuch [stw **734**], Frankfurt am Main; transl. as *The Law of Causality and its Limits* [The Vienna Circle Collection **22**], ed. by Robert Cohen, Kluwer, Dordrecht 1997.

Ginenthal, Charles (1995): *Carl Sagan & Immanuel Velikovsky*, New Falcon Publications, Tempe AZ.

Good, Irving John (1969): "A Subjective Evaluation of Bode's Law and an 'Objective' Test for Approximate Numerical Rationality", *Journal of the American Statistical Association* **64**, 23-66.

Hayes, Wayne, and Tremaine, Scott (1998): "Fitting Selected Random Planetary Systems to Titius–Bode Laws", *Icarus* **135**, 549-557.

Jeffrey, Richard (1965): *The Logic of Decision*, McGraw-Hill, New York.

Juergens, Ralph E. (1967): "Minds in Chaos", in Alfred de Grazia, Ralph Juergens and Livio Stecchini (eds.) (1967): *The Velikovsky Affair. Scientism vs. Science*, University Books, New Hyde Park NY.

Hacohen, Malachi (2000): *Karl Popper – The Formative Years, 1902–1945*, Cambridge University Press, Cambridge.

Hills, John G. (1969): "The origin and Dynamical Evolution of the Solar System", Ph.D. thesis, University of Michigan, Ann Arbor.

Hills, John G. (1970): "Dynamic Relaxation of Planetary Systems and Bode's Law", *Nature*, **225**, 840-842.

Holton, Gerald (1969): "Einstein, Michelson, and the 'Crucial' Experiment", *Isis*, **60**, 133-197; reprinted in Holton, Gerald (1973): *Thematic Origins of Scientific Thought. Kepler to Einstein*, Harvard University Press; 2nd ed. 1988.

Köhler, Eckehart (2001): "Why von Neumann Rejected Carnap's Dualism of Information Concepts", in Miklós Rédei and Michael Stöltzner (eds.) (2001): *John von Neumann and the Foundation of Quantum Physics* [Vienna Circle Institute Yearbook **8**], Kluwer, Dordrecht.

Köhler, Eckehart (2002): "Gödels Platonismus", in Bernd Buldt, Eckehart Köhler et al. (2002): *Kurt Gödel. Wahrheit & Beweisbarkeit*, öbv&htp, Vienna.

Köhler, Eckehart (2004): "Intuition Reclaimed. Gödel's Views, and How Carnap Abandoned Empiricism", forthcoming.

Kolmogorov, Andrey Nikolaevich (1954): "On the Conservation of Conditionally Periodic Motions for a Small Change in Hamilton's Function", *Doklady Akademsk Nauk SSSR* n.s. **98**, 527-530 (Russian); English transl. in Casatian & Ford (1979): *Stochastic Behavior in Classical and Quantum Hamiltonian Systems. Volta Memorial Conference, Como 1977* [Lecture Notes in Physics **93**], Springer-Verlag, Berlin.

Kuipers, Theo (1978): *Studies in Inductive Probability and Rational Expectation* [Synthese Library **123**], Kluwer, Dordrecht.

Langley, Pat, Herbert A. Simon, Gary L. Bradshaw and Jan M. Zytkov (1987): *Scientific Discovery. Computational Explorations of the Creative Processes*, MIT Press, Cambridge MA.

Laskar, Jacques (1994): "Large-Scale Chaos in the Solar System", *Astronomy and Astrophysics*, **287**, 9-12.

Lenat, Douglas (1982): "AM: Discovery in Mathematics as Heuristic Search", in Randall Davis and Douglas Lenat (1982): *Knowledge-Based Systems in Artificial Intelligence*, McGraw-Hill, New York.

Laskar, Jacques (2000): "On the Spacing of Planetary Systems", *Physical Review Letters*, **84**, 3240-3242.

Menger, Karl (1923): "Über die Dimensionalität von Punktmengen, I. Teil", *Monatshefte für Mathematik und Physik*, **33**, 148-160.

Menger, Karl (1925): "Grundzüge einer Theorie der Kurven", *Mathematische Annalen*, **95**, 277-306.

Moser, Jürgen (1973): *Stable and Random Motions in Dynamical Systems, with Special Emphasis on Celestial Mechanics* [Annals of Mathematics Studies **77**], Princeton University Press, Princeton NJ; reprinted 2001.

Murray, Carl (2002): "Is the Solar System Stable?", Fortune City, http://www.fortunecity.com/emachines/e11/86/solarsys.html.

Newell, Allan, and Simon, Herbert (1956): "The Logic Theory Machine – A Complex Information Processing System", *IRE Transactions on Information Theory*, IT-**2**, 61-79.

Newell, Allan, Shaw, J. Clifford, and Simon, Herbert (1962): "The Processes of Creative Thinking", in Gruber, Terrell & Wertheimer (eds.): *Contemporary approaches to Creative Thinking*, MIT Press; reprinted in Simon (1979).

Neyman, Jerzy, and Pearson, Egon S. (1933): "*On the Problem of the Most Efficient Tests of Statistical Hypotheses*", Philosophical Transactions of the Royal Society, **CCXXXI**, 289-337.

Nieto, Michael Martin (1972): *The Titius–Bode Law of Planetary Distances: Its History and Theory*, Pergamon Press, Oxford.

Nieto, Michael Martin (1974): "The Titius–Bode Law and the Evolution of the Solar System", *Pensée*, **4**, No. 3.

Niiniluoto, Ilkka, and Tuomela, Raimo (1973): *Theoretical Concepts and Hypothetico-Inductive Inference*, Reidel Publishing Co., Dordrecht.

Ovenden, Michael W. (1973): "Planetary Distances and the Missing Planet", *Recent Advances in Dynamical Astronomy*, Reidel Publishing Co., Dordrecht, 319-332.

Ovenden, Michael W. (1976): "The Principle of Least Interaction Action", in V. Szebehely & B.D. Tapley (eds.) (1976): *Longtime Predictions in Dynamics*, Reidel, Dordrecht, *pp.* 295-305.

Pais, Abraham (1982): *'Subtle Is the Lord ...'. The Science and Life of Albert Einstein*, Oxford University Press, Oxford.

Polya, George (1954): *Mathematics and Plausible Reasoning* **I:** *Induction and Analogy in Mathematics*; **II:** *Patterns of Plausible Inference*, Princeton University Press, Princeton.

Roy, Archie E., and Ovenden, Michael W. (1954): "On the Occurrence of Commensurable Mean Motions in the Solar System", *Monthly Notices of the Royal Astronomical Society*, **114**, 232-241.

Roy, Archie E. (ed.) (1991): *Predictability, Stability and Chaos in N-Body Dynamical Systems*, NATO ASI Series **B272**, Plenum, Oxford.

Roy, Archie. E. (1988): *Orbital Motion*, 3rd ed., The Institute of Physics, London.

Sagan, Carl (1974): "Venus and Dr. Velikovsky", in Carl Sagan (1979): *Broca's Brain*, Random House, New York.

Scheck, Florian (1999): *Mechanics – From Newton's Laws to Deterministic Chaos*, 3rd ed., Springer-Verlag, Berlin; 1st ed. 1990.

Scheck, Florian (2002): *Mechanik – Von den Newton'schen Gesetzen zum deterministischen Chaos*, 7th ed. Springer-Verlag, Berlin; 1st ed. 1988.

Siegel, Carl Ludwig, and Moser, Jürgen K. (1971): *Lectures on Celestial Mechanics*, Springer-Verlag, Berlin.

Simon, Herbert (1977): *Models of Discovery and Other Topics in the Methods of Science* [Synthese Library **114**], Reidel Publishing Co., Dordrecht.

Simon, Herbert (1979): *Models of Thought*, Yale University Press, New Haven CT.

Spohn, Wolfgang (1991): "A Reason for Explanation: Explanations Provide Stable Reasons", in Wolfgang Spohn, Bas C. van Fraassen and Brian Skyrms (eds.) (1991): *Existence and Explanation*, Kluwer, Dordrecht.

Spohn, Wolfgang (1993): "Causal Laws Are Objectifications of Inductive Schemes", in Jacques-Paul Dubucs (ed.) (1993): *Philosophy of Probability*, Kluwer, Dordrecht.

Spohn, Wolfgang (2002) "Laws, Ceteris Paribus Conditions and the Dynamics of Belief", *Erkenntnis* **57**, 373-394.

Sussman, Gerald Jay, and Wisdom, Jack (2001): *Structure and Interpretation of Classical Mechanics*, MIT Press, Cambridge MA.

Velikovsky, Immanuel (1950): *Worlds in Collision*, Macmillan (later the same year Doubleday), New York.

Wilson, Fred (1997): *Hume's Defense of Causal Inference*, University of Toronto Press, Toronto.

Wisdom, Jack (1981): "Chaotic Dynamics in the Solar System", *Proceedings of the Royal Society* **413**, 102-129; reprinted in *Icarus*, **72**, 241-274.

Dept. of Business Administration
University of Vienna
Austria
eckehart.koehler@univie.ac.at

PETER CLARK

FREGE, NEO-LOGICISM AND APPLIED MATHEMATICS

I. INTRODUCTION – LOGICISM AND NEO-LOGICISM

A little over one hundred years ago (the letter is dated July 28, 1902), Frege wrote to Russell in the following terms[1]:

> I myself was long reluctant to recognize ranges of values and hence classes; but I saw no other possibility of placing arithmetic on a logical foundation. But the question is how do we apprehend logical objects? And I have found no other answer to it than this, We apprehend them as extensions of concepts, or more generally, as ranges of values of functions. I have always been aware that there are difficulties connected with this, and your discovery of the contradiction has added to them; but what other way is there?

Frege here poses an extremely good question, a recent answer to which this paper is really devoted. Whatever view one may finally adopt about whether the new answer succeeds, one has to recognize that it is surely a very remarkable fact that one hundred years after the discovery of the Zermelo-Russell contradiction, which follows from Basic Law Five of Frege's *Grundgesetze*, we should now be actively discussing, not as a purely historical enterprise, but as a viable possibility in the foundations of mathematics, the programme more or less explicitly laid out in the earlier work of Frege *Die Grundlagen der Arithmetik*. This is because recent research has highlighted three crucial facts. First and most importantly, that in full conformity with the spirit of Frege's programme, the deduction of the axioms of (Second Order) Peano Arithmetic from principles of higher order logic and 'definition'[2] does not require appeal to Basic Law Five. Second, that the principles of logic and 'definition' (this definition has become known as Hume's Principle) which are employed are consistent[3], and, third, that even the proofs of the axioms for Peano Arithmetic, given by Frege himself in the *Grundgesetze der Arithmetik*, do not depend essentially upon Basic Law Five[4]. These technical facts, of which more later, open the way for the revival of Frege's programme. They make it possible, but they do not determine its form.

What was Frege's programme and what is its revived form explicitly? Frege's programme was the result of an answer to the famous question raised in Section 62 of the *Grundlagen* viz.: 'how then are numbers given to us, if we cannot have any ideas or intuitions of them?' The Fregean answer was 'by explaining the senses of identity statements in which number words occur'. That

F. Stadler (ed.), Induction and Deduction in the Sciences, 169–183.
© 2004 *Kluwer Academic Publishers.*

explanation was to be provided at least in part by what has come to be called Hume's Principle: the claim that the cardinal numbers corresponding to two concepts are identical if and only if the two concepts are equinumerous. I say in part by Hume's Principle because as Frege had already argued in another context at Section 56 of the Grundlagen whatever the merits of Hume's Principle, it cannot explain the senses of identity statements in which number words occur of the form the number of F's is n, where n is not given in the form of 'the number of G's, for some G. Frege then adopted the explicit definition of number in terms of classes or extensions 'the number of F's is the class of all concepts G, equinumerous with F'. But this explicit definition together with Basic Law Five, the comprehension axiom for class existence, entails Hume's Principle. With Axiom Five in place, it looked as if Frege's programme could be carried out.

It was now possible to show that second order logic together with Basic Law Five entails the Peano Dedekind axioms for arithmetic. As such, the truths of arithmetic could be seen to be analytic, and could all be seen to be consequences of general logical laws together with suitable implicit definitions (like Basic Law Five which implicitly defines the notion of an extension). Further, arithmetic could be seen as a body of truths about independently existing objects – the finite cardinals – which were logical objects, logical in the sense that knowledge of which requires nothing beyond knowledge of logic and definitions. If this could be extended to the Real numbers and other parts of mathematics, then a foundation for mathematics would have been established on which mathematics was presented as uncontaminated by empirical notions, presented as a body of truths in its full classical form and shown to be applicable to reality, since it can be shown to be in fact the more and more elaborate drawing of consequences from meaning postulates. But deduction applies to everything that can be thought. In fact, it seems to me that the fundamental logicist thought can be put simply as follows: there can be no thought without representation, there can be no representation without concepts, and there can be no concepts without number.

Of course the serpent had already entered this Eden, with the introduction of Basic Law Five which says

$$(\forall F)(\forall G)(\, \mathrm{Ext}\,(F) = \mathrm{Ext}\,(G) \leftrightarrow (\forall x)(Fx \leftrightarrow Gx))$$

But by the Comprehension Principle for Second Order Logic, there is a property corresponding to the formula of Second order logic $(\exists F)(\mathrm{Ext}\,(F)=x\ \&\neg Fx)$. Russell's paradox immediately results from allowing this property to fall under the universal quantifier $(\forall F)$ in Basic Law Five. Another way of putting the same point is to note that the Russell reasoning shows that it is a theorem of Second order logic that there is no function from properties to objects such that distinct properties (i.e., non-coextensive properties) are associated with distinct objects. This is just what Basic Law Five, read from left to right in contrapositive form, asserts there is[5].

So much for logicism. What about Neo-logicism, the revived form of Frege's programme? We should let Wright and Hale, the main proponents of this view speak for themselves. They say[6]:

> Neo-Fregeanism holds that Frege need not have taken the step which lead to this unhappy conclusion [The appearance of the Russell contradiction]. At least as far as the theory of natural number goes, it is possible to accomplish Frege's central mathematical and philosophical aims by basing the theory on Hume's Principle, adjoined as a supplementary axiom to a suitable formulation of second order logic. Hume's Principle cannot, to be sure, be taken as a definition in any strict sense – any sense requiring that it provide for the eliminative paraphrase of its definiendum (the numerical operator, "the number of...") in every admissible type of occurrence. But this does not preclude its being viewed as an implicit definition, effecting an introduction of a sortal concept of cardinal number and, accordingly, as being analytic of the concept – and this, the neo-Fregean contends, coupled with the fact that Hume's Principle so conceived requires a prior understanding only of second order logical vocabulary, is enough to sustain an account of the foundations of Arithmetic that deserved to be viewed as a form of logicism which, whilst not quite logicism in the sense of a reduction of arithmetic to logic, preserves the essential core and content of Frege's two fundamental theses.

So two conditions will have to hold for this position to viable: first, what has come to be known as Frege's theorem (which certainly does hold), that is, the mathematical claim that the Dedekind-Peano axioms for number theory in their second order form can be derived from a combination of Second order logic and Hume's Principle, and second, that it will have to be shown that Hume's Principle and other so-called abstraction principles, which share its form, constitute a legitimate means of introducing the names of numbers (of abstract objects in general) by, in effect, stipulation by implicit definition. One needs to be careful to state the claim of neologicism properly, it is:

(i) Hume's Principle is a stipulation which gives the truth conditions of a restricted class of statements of numerical identity

(ii) The resulting explanation of the concept of number is complete, however, in that it suffices for the second order derivation of the basic laws of arithmetic

(iii) The existence of numbers is something discovered and not stipulated (the Platonism of Frege's original theory is preserved)

(iv) Our (a priori) *knowledge* of number is derived from a principle whose truth is a matter of stipulation[7].

Abstraction principles, of which Hume's Principle is a paradigm example, come in two types conceptual abstractions and objectual ones, but all have the following form. There is a domain of entities, denoted say, by α, β, etc., and a relation R defined over them. Then an abstraction principle has the form

$$(\Sigma(\alpha) = \Sigma(\beta)) \leftrightarrow R(\alpha,\beta)$$

where R(,) is an equivalence relation among the α and β's. An abstraction principle may be called a logical abstraction when the relation R(,) is definable in purely logical vocabulary, e.g., equinumerosity among concepts or ordinal similarity among binary relations. Under the classical canonical interpretation, $\Sigma(\alpha)$ is the equivalence class of α under the relation R, and exists (where it does) in virtue of a set existence axiom. That is, the existence and uniqueness of $\Sigma(\alpha)$ has in effect to be guaranteed by a separate principle of set or class existence. Wright and Hale, however, argue that in certain cases logical abstraction principles can play the role of stipulations and if the relation on the right hand side of the iff is ever satisfied, then no further question concerning the existence of the $\Sigma(\alpha)$ need arise. Conceptual abstraction principles are those in which α's are concepts (as in the case of Hume's Principle), and objectual abstraction principles are those in which the field of the equivalence relation comprises objects. In both cases, and this is crucially so, the abstracta the $\Sigma(\alpha)$ are objects, so in the case of conceptual abstractions Σ acts as a type down operation, from concepts to objects.

Of course Wright and Hale do not argue that it is always legitimate to introduce abstracta in this way. Two examples of conceptual logical abstraction principles which fail to introduce abstracta are Basic Law Five and what might be called Ordinal Hume which is the claim that

$$(\forall R)(\forall S)(\text{Ord } R = \text{Ord } S \leftrightarrow R \text{ is similar to } S).$$

This has the form of an abstraction principle since similarity is an equivalence relation among binary relations. But Ordinal Hume leads directly to the Burali-Forti paradox[8]. Hume's Principle in logical abstraction form says:

$$(\forall F)(\forall G)(\ NxFx = NxGx \leftrightarrow\ (\exists R)(F\ \ 1\text{-}1_R\ G))$$

where $(\exists R)(F\ \ 1\text{-}1_R\ G)$ is an abbreviation for the standard formulation in the vocabulary of second order logic of the formula expressing that there is a relation R which establishes a one to one correspondence between the things falling under F and those falling under G, that is

$$((\exists R)(\forall x)(Fx \rightarrow (\exists!y)(Gy\ \&\ R(x,y))\ \&\ \forall z(Gz \rightarrow (\exists!w)(Fw\ \&\ R(w,z)))$$

and the operator Nx...x is a term forming operator. Wright has argued that there are general principles which can distinguish between good and bad abstraction principles, and in any case, there is no similar problem about Hume's Principle, since it is known to be consistent. Like Basic Law Five, Hume's Principle asserts the existence of a function from concepts to objects, but unlike Basic Law Five it asserts that merely non-equinumerous concepts (not non co-extensive concepts) can be sent to distinct objects and this is possible provided that the domain is (Dedekind) infinite. For a domain of k objects, there are k+1 non-equinumerous concepts definable over it, so no finite domain can satisfy Hume's Principle.

The values of the second order variables for Frege are concepts and objects denoted by terms that may appear on either side of the identity sign, so the terms like NxFx denote objects. As such, Nx x will have to be thought of as a term forming operator and our theory of second order logic plus Hume's Principle must be sufficiently strong to have as a theorem $(\forall F)(\exists!x)(x=NxFx)$. In general, one would expect that we would have to have an axiom asserting that for each F, NxFx was a term.

This fact seems to me to have the profoundest significance for the claim that Hume's Principle and other abstraction principles can be thought of as analytic stipulations introducing the names of numbers (or other abstract objects). Nor do I see any of the various equivalent methods of adding Hume's Principle to second order logic in order to derive Frege's theorem as avoiding this difficulty, for they will all have to guarantee that $(\forall F)(\exists!x)(x=NxFx)$ holds in some way or other. If it is guaranteed, then we may proceed as follows: consider the concept non-self identical. It is a truth of logic that the concept non-self identical is equinumerous with the concept non-self identical, so the right hand side of Hume's principle for F and G both non-self identical reduces to a logical truth. So we may detach and assert as a theorem $Nx\sim(x=x) =Nx\sim(x=x)$ and so infer $(\exists!y)(Nx \sim(x=x) =y)$ and so introduce the name 'zero' to designate that unique object. But that hinges on us being able to read the stipulation, to understand the stipulation, as providing a context into which it is appropriate to quantify in. But what guarantees that that is the correct reading of the stipulation? This is my first sceptical argument then: it essentially concerns how the underlying logic is to function. On the one hand, to presume NxFx is a term because it appears on one side of the identity sign seems to me to beg the question, while if the question is not begged, and say a free logic is employed, then I fail to see how the required existential postulates will form any intended contrast with "mere axiomatic postulation"[9]. But I will not dwell further on this matter here.

II. ANTI-ZERO

Following Boolos we can write Hume's Principle in the form

$$*F=*G \leftrightarrow (F \text{ 1-1 } G)$$

and understand it as asserting that there exists a total function from concepts to objects, call it *, such that non-equinumerous concepts are assigned distinct objects (that is the contrapositive of Hume's principle read from left to right). Adding this principle to Second order logic allows us to prove in the Fregean way that with the successor relation defined in the usual manner viz.

$$S(n,m) \leftrightarrow(\exists F)(\exists y)(Fy \,\&*F =n \,\&*(Fx \,\&x \neq y)=m)$$

(i) the successor relation is functional and one-one (ii) but that with zero defined as *(x ≠ x), (∀x)~S(0,x) and n is Finite iff n is zero or S^(n,0) where S^ is the ancestral relation of S. But that is just to say that the natural numbers form a Dedekind infinite sequence, that is:

$$(\exists f)[(\forall x)(\forall y)(fx = fy \rightarrow x=y)\&(\exists x)(\forall y)(fy \neq x)].$$

The key step is clearly to be able to prove that every number has a successor and Frege's proof works precisely because n is an object that can be proved not to fall under the concept 'being less than n'. This was one of Frege's triumphs in the *Grundlagen*. It might, therefore, seem little more than trite to claim that H.P. is an axiom of infinity. It clearly is, since it forces the domain to be infinite. No doubt it is, but the principle has many other odd entailments too. Some are surprisingly and worryingly strong.

Since the function * is total, we can take it that it is a theorem of Frege Arithmetic that (∀F)(∃!x)(x = *F). Call this theorem (**). Now, as Frege noted, if we let F be "... is a member of the natural sequence of numbers", it follows that there is a number of finite numbers, Frege's ∞_1, and that since it succeeds itself, it is not a finite number (since it is a theorem of the formal system implicit in the *Grundlagen* that if Finite(n) then ~S(n,n)). Again, it follows immediately from (**) that if we take F to be the concept "is self identical", then (∃!x)(x = *(x =x)), that is, that there is a number of all things that there are. This latter looks like a very strong claim indeed. But we should note immediately, as Boolos pointed out, that we cannot show within the theory (Frege Arithmetic) that the two numbers are distinct[10]. In what we might call the standard model of Hume's Principle, with domain N, the natural numbers, in which all infinite concepts are assigned the object zero, while all finite concepts are assigned the cardinality of the corresponding subset of N plus one, the object associated with *Finite (n) is zero, as is the object associated with *(x=x), since neither are finite. But the set of all ordinals less than or equal to Aleph one is also a model for the theory (that is for Frege Arithmetic), and in this case *Finite (n) is assigned Aleph zero while *(x=x) is assigned Aleph one. So we certainly cannot prove in Frege Arithmetic that *Finite (n) ≠ *(x=x).

But a worry still remains, for as Boolos asked "is there such a number as anti-zero?"[11] That seems to me a very good point indeed. For that, if anything is, seems to be a substantial matter, which cannot be decided by stipulation. Indeed, this issue seems to generalise as Boolos pointed out into a general issue about the compatibility of two conceptions of cardinal number. One derived from what one might call the pure theory of cardinal number based on Hume's Principle and one derived from set theory as we shall see below.

However, it is possible to dissolve this worry and Wright has done so. He points out that it was a prime tenet of Frege's view that numbers were the numbers of *sortal* concepts. Clearly, the concept "self identical" is not a genuinely

sortal concept, so it is not an appropriate instance of the theorem above (**). Thus, as Wright remarks in his reply to Boolos:

> Moral: just as not every object is suitable to determine a direction, so we should not assume without further ado that every concept – every entity an expression for which is an admissible substituted for the bound occurrences of the predicate letters in Hume's Principle – is such as to determine a number.

He goes on to say :

> So self-identity is not a sortal concept. If we take it that, save where F is assured an empty extension on purely logical grounds, only sortal concepts, and concepts formed by restricting a mere predicable to a sortal concept, have cardinal numbers, it follows that there is no universal number.[12]

So the function * is not to be regarded as having as its domain all concepts, rather there is to be natural and principled restriction on its domain of interpretation. It is to be confined to sortal concepts. However, it is now unclear whether we can any longer regard Hume's Principle as a purely logical abstraction principle, since it is not clear that the notion of being a sortal concept can be expressed in purely logical vocabulary. But, as Wright was quick to note, if we do accept the need to restrict the domain of the function * to genuinely sortal concepts, then a second difficulty looms and to some of us it looms very large.

III. THE GOOD COMPANY OBJECTION

Let us follow Wright's lead and consider the sortal concepts "set", "ordinal" and "cardinal number". According to Hume's Principle, these concepts too ought to have a number associated with them by *. But then we immediately invite the observation that there is bound to be conflict with the notion of number as embodied in set theory, say ZFC. According to ZFC, there are no numbers associated with the concepts "set", "ordinal" and "Cardinal number" precisely because the extensions of such concepts do not form sets and ZFC embodies the principle: no set no cardinal number. The collections we have been considering are proper classes, not sets, so there is no number associated by ZFC (or set theory) with them. Considerations such as these suggested to Boolos that there were two conflicting conceptions of number in play. He remarks[13]:

> Two thoughts about the concept of number are incompatible: that any zero or more things have a (cardinal) number, and that any zero or more things have a number (if and) only if they are the members of some one set. It is Russell's paradox that shows the thoughts incompatible: the sets that are not members of themselves cannot be the members of any one set. The thought that any (zero or more) things have a number is Frege's; the thought that things have a number only if they are the members of a set may be Cantor's and is in any case a commonplace of the usual contemporary presentations of the set theory that originated with Cantor and has become ZFC.

In a similar vein, he elabourated on the issue[14]:

> The worry is that the theory of number we have been considering, Frege Arithmetic, is incompatible with Zermelo-Fraenkel set theory plus standard definitions, on the usual and natural readings of the non-logical expressions of both theories. To be sure, as Hodes once observed in conversation, if *a is taken to denote the cardinal number of a when a is a set and some favourite object that is not a cardinal number when a is a proper class, then HP will be a theorem of Von Neumann set theory. But on that definition of *, * will not be translatable as 'the cardinal number of.' ZF and Frege arithmetic make incompatible assertions concerning what cardinal numbers there are. And of course, the response 'Well, these are just formalisms; the question of their truth or falsity doesn't arise or makes no sense' is hardly available to one claiming that HP is analytic, i.e., an analytic *truth*. So one who seriously believes that has to be bothered by the incompatibility of the consequence of Frege arithmetic that there is such a number as anti-zero with the claim made by ZF + standard definitions (on the natural reading of its primitives) that there is no such number.

Since to be in the company of set theory is to be in very good company indeed, let us call this objection to the stipulatory nature of Hume's Principle, the Good Company objection. Now, there is a clear response which can be made by someone who wishes to defend the idea that the truth of Hume's principle can be simply stipulated, and that is that the conflict alluded to above is illusory.[15] After all set theory assigns numbers to sets, but Frege arithmetic assigns numbers to concepts. Frege arithmetic assigns a cardinal number to the Russell concept "non-self membered set", but in virtue of the way cardinal numbers are introduced in ZFC, no cardinal number can be assigned to a non-set. (It is worth recalling that in ZFC if X is any set, then there is an ordinal number α and a bijection $f:\alpha\rightarrow X$. For any set X the cardinality of X ($|X|$) is the least ordinal α such that there is a bijection $f:\alpha\rightarrow X$. A cardinal number is an ordinal number α such that for no $\beta<\alpha$ is there a bijection $f:\beta\rightarrow\alpha$.) The response to the Good Company objection in short, then, is Frege arithmetic assigns a number to proper classes, ZFC is silent. In any case, certainly, there is no conflict. But this dismissal of the good company objection is too swift.

Let us go back to the first quotation from Boolos. He rightly points out that two *conceptions* of number are incompatible. The first conception is embodied in Hume's Principle and says that every sortal concept has a number and the other says that every sortal concept has a number if and only if the extension of that concept is a set. Call these the Fregean and Non Fregean conceptions, respectively. Clearly, Russell's paradox does indeed show that these two conceptions are incompatible. "Non self membered set" is certainly a sortal concept and so has a number, by the Fregean conception. But it is provably the case that the concept "non self membered set" has no set, as its extension so on the second conception by the contrapositive of the only if clause there can be no number corresponding to the concept.

Now imagine someone, we had better call him "Anti-Hero" or simply "Villain" (in this context), who believes this: the conception of number we have is revealed in mathematical practice and he holds to the Non-Fregean concept of

number. Is it conceivable that he holds an analytically false belief, or a belief that a stipulation may show to be false? Our grasp of number is what we do with it in mathematics. Now, it is certainly true that the Non Fregean concept of number outlined above is entirely adequate for the whole of mathematics, since it is realised by ZFC (though ZFC does not entail it). The extra content which Anti-Hero in fact believes over ZFC is the claim that the mathematical universe is exhausted by ZFC. That claim certainly seems believable. What it certainly does not seem to be is analytically false and it seems to have enormous evidence in favour of it. It may seem that Anti-Hero is merely legislating on the content of all possible mathematics and that cannot be right. But Anti-Hero doesn't believe that his view is analytically true; what he certainly believes is that it is not analytically false however.

Perhaps it was something like this that Boolos had in mind when he wrote of Anti-Hero's view that: "It is in any case a commonplace of the usual contemporary presentations of the set theory that originated with Cantor and has become ZFC". Clearly, this matter must be connected with Boolos's rejection of the notion of proper class. He took the view that they are in fact just "a manner of speaking", in the sense that they really play the role of abbreviations whose use can always be eliminated in any formal theory by replacing them by their defining formulas. This latter view I think is directly connected with his support for the iterative conception of set and the claim that ZFC exhausts that iterative hierarchy. Whatever may be the case about this, it does seem to me that Boolos's Good Company objection is a compelling one, for the set theoretic conception of number is a perfectly viable one and it surely cannot be rejected as analytically false. A possible response would be to say that there are two conceptions of cardinal number one for concepts and one for sets, but then what would have happened to the Fregean foundational programme that this was the correct ontological and epistemological account of the nature of and of our knowledge of *the* cardinal numbers? We would then seem to be left with an impenetrable problem about reference. We have an account of Frege numbers, and we have an account of 'set' numbers, but neither is apparently related to each other or the numbers of ordinary arithmetic. That response would amount to abandoning the foundational programme and would hardly be acceptable to the Neo-Fregean.

IV. INDEFINITELY EXTENSIBLE CONCEPTS AGAIN-PROPER CLASSES

Perhaps it was some considerations such as these which lead Wright to endorse the principle at the heart of the non-Fregean concept of number as characterised above. He writes[16]:

> Grant the plausible principle that there is a determinate number of F's just provided the F's compose a set

and later

... it seems natural and well-motivated to suppose that the F's should have a determinate cardinal number just when they compose a set.

If this is the line to follow, then what we are looking for is some further principled restriction on range of the quantifiers in Hume's Principle over and above that of the restriction to sortal concepts. Wright proposes that the concepts falling under the range of the universal quantifiers in Hume's principle should be restricted to those which are 'definite'. Recall that a concept is definite just when it is not "indefinitely extensible".

One sense of indefinite extensibility goes back to Russell and Poincaré. Russell, in 1903, had thought that the contradiction, derivable from Basic Law V of Frege's *Grundgesetze*, showed that not every property determines a class *simpliciter*.[17] The fundamental question as he saw it was then 'to determine, which propositional functions define classes which are single terms as well as many, and which do not?' By 1906, after reading Poincaré, he had changed his view. He wrote 'the contradictions result from the fact that, according to current logical assumptions, there are what we may call self-reproductive processes and classes. That is, there are some properties such that, given any class of terms all having such a property, we can always define a new term also having the property in question'[18]. In formulating the Vicious Circle Principle, he made a very similar claim: 'Thus all our contradictions have in common the assumption of a totality such that, if it were legitimate, it would at once be enlarged by new members defined in terms of itself'[19]. Both the Russell of this period and Poincaré would have agreed that the objects falling under such self-reproductive properties like ordinal or set form no totality. As is very well known, Dummett has recently revived this idea[20].

As Wright remarks of the idea[21]:

I do not, myself, know how best to sharpen this idea, still less how its best account might show that Dummett is right both to suggest that the proof-theory of quantification over indefinitely extensible totalities should be uniformly intuitionistic and that the fundamental classical mathematical domains, like those of the natural numbers, or the reals, should also be regarded as indefinitely extensible. But Dummett could be wrong about both those points and still be emphasizing an important insight concerning certain very large totalities – ordinal number, cardinal number, set, and indeed 'absolutely everything'.

So the idea that Wright is proposing, then, is this I take it. We should restrict the range of the quantifiers in Hume's principle to definite concepts, that is, those that are not indefinitely extensible. But one has to careful here, for there are at least two conceptions of indefinite extensibility, if there are any; one which is very closely related to the Russell-Wright line of thought which sees indefinite extensibility associated with concepts generating or holding of very large totalities, with in effect, proper classes (in the presence of strong versions of the axiom of choice all the proper classes we have been considering are equinumerous with the Universe), and one which Dummett has developed in which both N and R are indefinitely extensible. Deploying Dummett's conception will of

course have essentially the same effect as deploying a suggestion due to Heck[22] to the effect that analytic core of Hume's Principle is Finite Hume. This is the Principle that

$$(\forall F)(\forall G)(\text{Fin}(F) \,\&\, \text{Fin}(G) \rightarrow (\, {}^*F = {}^*G \leftrightarrow (F \text{ 1-1 } G))$$

where the notion of finite can be spelled out in purely logical vocabulary (as the negation of the second order sentence expressing that F is (Dedekind) infinite). This is, of course, not in the form of an abstraction principle and so is unacceptable as an implicit definition for the Neo-Fregean. The closely related principle

$$(\forall F)(\forall G)((\, {}^*F = {}^*G \leftrightarrow \text{Not Fin}(F) \text{ V Not Fin}(G) \text{ V } (F \text{ 1-1 } G))$$

although it is an abstraction principle (the right hand side being an equivalence relation), would not suffice either, since it conflicts with the Cantorian conception of cardinal number by assigning the same number to all infinite concepts.

However what would seem to capture Wright's restriction is precisely the notion of set as opposed to proper class. I take it what we want to say is that N and R are definite while "set", "ordinal" and "cardinal number" are not. Can such a notion of definite be made out? Well, of course it can: the natural candidate is, a concept F is definite iff it has a set as its extension. This will do exactly what Wright wants. But it would make the understanding of Hume's principle parasitic upon the notion of set and our grasp of the set theoretic Universe. It could hardly then be argued that the truth of Hume's Principle was guaranteed by stipulation.

We have already noted that in the presence of a very strong choice principle, the concepts we wish to exclude from the domain of Hume's Principle are equinumerous with the Universe. We could use this fact to independently give a justification of Wright's New Hume which would have the form

$$(\forall F)(\forall G)(\, {}^*F = {}^*G \leftrightarrow (\text{InDef } (F) \,\&\, \text{InDef}(G)) \text{ v } (F \text{ 1-1 } G))$$

where InDef (F) would be a condition stating that F is equinumerous with the Universe.

This would still be a logical abstraction principle, provided that the condition InDef (F) were expressed by the formula:

$$(\exists R)((\forall x)(Fx \rightarrow (\exists !y)Rxy) \,\&\, (\forall y)(\, \exists !x)(Fx \,\&\, Rxy))^{23}$$

However one proceeds, here there looks to be a dilemma: either restrictions on the range of the quantifiers in Hume's principle or (*) are actually motivated by set theoretic considerations, or some notion of definiteness can really be made out, which itself does not rely upon set existence principles. These would serve as a principled restriction, but in either case clearly something other than the no-

tion of (sortal) concept and of second order logic is motivating our understanding of Hume's Principle, and thus of our knowledge of number. Whatever route is the correct one, it seems hardly possible to regard the laying down of such principles as guaranteeing their truth by stipulation.

V. FREGE AND THE APPLICATION OF ARITHMETIC

It is certainly true that Frege put the application problem at the heart of his philosophy of mathematics. To the question: 'why does arithmetic apply to reality?', the logicist provides the clear answer, because it applies to everything that can be thought. It is the most general science possible. The partial contextual definition, provided by Hume's Principle and the fundamental thought that numerical concepts are second level concepts, yield Frege's account of the applicability of mathematics. In the simplest case for which the question arises – the application of the cardinal numbers – the solution is that arithmetic is applicable to reality because the concepts, under which things fall, themselves fall under numerical concepts. Thus, it is possible to prove in second order logic that $\exists^n xFx$ – F falls under the numerical property expressed by the numerically definite quantifier $\exists^n x$ if and only if the Frege numeral introduced by the partial contextual definition (Hume's Principle) is indeed n. In other words, the theorem that $\exists^n xFx \equiv n = NxFx$ can be obtained from Hume's Principle in second order logic. But there is a real difficulty with Frege's solution to the application problem, and that is that we are provided by Hume's principle with at best a partial contextual definition. The principle cannot settle the truth conditions of sentences of the form q = NxFx where q is not given in the form NxHx for some H. This, of course, is the famous Julius Caesar problem. In the case of pure arithmetic, the Julius Caesar problem can properly be regarded as irrelevant, since there would be no other singular terms in the language, but once the language is extended to include empirical singular terms, as it would in the language of applied arithmetic, Hume's Principle will no longer settle the sense of numerical identities and the solution to the application problem will fail. Of course, this issue does not arise in the formal language of the *Grundgesetze*, which is a language of pure arithmetic, since in that language all the objects it is possible to refer to are already given, as extensions (value ranges), the identity conditions for which are given by Basic Law Five. But as Michael Potter has put it[24]: 'a formal language in which Julius Caesar cannot be spoken of is one in which he cannot be counted, and in such a language the applicability of arithmetic remains unexplained. At some stage in the development, we shall have to extend the formal language by adding some empirical vocabulary, and we shall then have to address the Julius Caesar problem just as before.'

The question naturally arises as to whether the neo-logicist fares any better than Frege with respect to the application problem. This seems very unlikely since the neo-logicist[25] relies exclusively upon Hume's Principle, and therein

lays the real difficulty, as Frege long ago knew. In the same letter which I quoted at the beginning of this paper, he wrote to Russell about the idea of letting his programme rest on Hume's Principle alone the following[26]:

We can also try the following expedient, and I hinted at this in my *Foundations of Arithmetic*. If we have a relation $\Phi(\xi,\zeta)$ for which the following propositions hold: (1) from $\Phi(a,b)$ we can infer $\Phi(b,a)$, and (2) from $\Phi(a,b)$ and $\Phi(b,c)$ we can infer $\Phi(a,c)$; then this relation can be transformed into an equality (identity), and can be replaced by writing, e.g., '§a = §b'. If the relation is, e.g., that of geometrical similarity, then 'a is similar to b' can be replaced by saying 'the shape of a is the same as the shape of b'. This is perhaps what you call 'definition by abstraction'. But the difficulties here are not the same as in transforming the generality of an identity into an identity of range of values.

NOTES

1. Frege (1902), pp. 140-41.
2. Actually, one has to be careful how this result is stated. Formally, the central result is that if a formalisation of the key 'definition' is added as an axiom to standard axiomatic second order logic, second order arithmetic (arithmetic with the full second order induction axiom) can be interpreted in the resulting theory, often called Frege Arithmetic. Certainly, the result seems to have been known to Geach in the forties, Dummett in the fifties and was first recently explicitly noted by Charles Parsons in his 1965 paper 'Frege's Theory of Number' (see Demopoulos (1995), pp. 182-210). A very closely related result was published by Timothy Smiley in 1981 (Smiley (1981)). The full significance of the result, as well as a well developed proof, was given by Crispin Wright in 1983 (Wright(1983)). The result has been systematically investigated by George Boolos (who discoved it independently in the early eighties), and Richard Heck. (See especially Boolos (1998), papers 17, 18 and 19, and Heck (1993). The most accessible proof can be found in the Appendix to Boolos (1990a), Boolos (1998) paper 13.)
3. Wright conjectured that the system of axiomatic second order logic together with Hume's Principle is consistent, but did not establish it. Burgess, Hazen and Hodes provided elementary consistency proofs with ω and $\omega+1$ as domains. George Boolos, however, established the central consistency result, which is that the theory is equi-consistent with analysis (see his (1987b) and Boolos and Heck, paper 20 of Boolos (1998)).
4. This was certainly known to Frege (see especially Heck (1995)), but in a letter to Russell he dismissed the possibility of basing his system on Hume's Principle, saying only that it faced difficulties which were different from those facing the attempt to use Basic Law Five. (Frege (1902) letter xxxvi/7, p. 141.)
5. This is Frege's own generalisation of the lesson of Russell's paradox.
6. Wright and Hale (2000), Introduction.
7. See also Demopoulos (1998), (2000).
8. See Hodes (1984).
9. Wright and Hale consistently draw attention to the difference in methodology they see between mere axiomatic postulation and their proposed methodology, whereby abstracta are taken as stipulations. I have not dwelt on what I regard in this respect as a separate issue, namely that of the very strong existential import of second order logic. This is most clearly seen in a point that George Boolos constantly emphasized that is the very strong existential commitments embodied in the Second Order Comprehension Principle $(\exists X)(\forall x)(Xx\leftrightarrow A(x))$ where $A(x)$ is a formula of second order logic not containing X free. The issue of the existential presuppositions of the Hume's principle has been addressed by Shapiro and Weir (2000) and by Demopoulos in his

review of M. Schirn: *Philosophy of Mathematics Today* in the Journal of Symbolic Logic. Shapiro and Weir conclude that 'the neologicist has no non-question begging account of how there could be an epistemologically innocent route to the demonstration of platonistically construed mathematical existence claims' (p. 188).

10. Boolos (1987b), p. 197 (page reference to the reprint in Boolos (1998)).
11. Boolos (1997), p. 314 (page reference to the reprint in Boolos (1998)).
12. Wright (2000) 'Is Hume's principle Analytic?' in the *Notre Dame Journal of Formal Logic*.
13. Boolos (1995) 'Frege's Theorem and the Peano Postulates', in Boolos (1998), p. 291.
14. Boolos (1997), p. 314. Page reference is to Boolos (1998).
15. This objection was put to me by Stewart Shapiro and Fraser MacBride.
16. Wright (2000) 'Is Hume's principle Analytic?' *Notre Dame Journal of Formal Logic*, forthcoming.
17. Russell (1903) *The Principles of Mathematics*.
18. Russell (1906) 'On some difficulties in the theory of transfinite numbers and order types', reprinted in *Bertrand Russell Essays in Analysis*, edited by D. Lackey London, George Allen and Unwin, 1973, pp. 135-64.
19. Russell (1908) 'Mathematical logic as based on the theory of types', reprinted in *Logic and knowledge*, edited by R.C. Marsh, London, George Allen and Unwin, 1956, pp. 59-102.
20. Dummett (1991) *Frege: Philosophy of Mathematics*.
21. Wright (2000) 'Is Hume's principle Analytic?' *Notre Dame Journal of Formal Logic*.
22. Heck (1997c).
23. Boolos suggested that we might regard the principle
    $(\forall F)(\forall G)(EXT(F)=EXT(G)\leftrightarrow(InDef(F)\&InDef(G))\ v\ (\forall x)(Fx\times Gx))$
    as a repair of Basic Law Five. This might well then be used as a justification of set theory based upon the Principle of the Limitation of size. But like that justification, however, the power set axiom might well prove a difficulty when taken together with the separation schema. Knowing that, for example, N is not "too big" does not help us with the claim that P(N) is not too big. Similarly, knowing that X is definite might tell us very little about or nothing at all about the definiteness of P(X). The effect of introducing New Five has been very carefully studied by Shapiro and Weir. (See Shapiro and Weir (1999).)
24. Potter (2000), p. 108.
25. Hale and Wright treat the Julius Caeser problem very seriously indeed. They say of it that it is 'one of the hardest the neo-Fregen must solve' (Hale and Wright (2001), pp. 14-16). They devote the whole of essay 14 of their ((2001), pp. 335-396) to the topic.
26. Frege (1902), p. 141.

REFERENCES

Boolos, G. (1987a), "Saving Frege from Contradiction" *Proceedings of the Aristotelian Society* 87 (1987), pp. 137-51.

Boolos, G. (1987b), "The Consistency of Frege's Foundations of Arithmetic" in Thomson (1987), pp. 3-20.

Boolos, G. (1990a), "The standard of equality of numbers" in Boolos (1990b), pp. 261-77.

Boolos, G. (Ed.) (1990b), *Meaning and Method: Essays in Honor of Hilary Putnam* (Cambridge: Cambridge University Press 1990).

Boolos, G. (1993), "Whence the Contradiction?" *Proceedings of the Aristotelian Society, Supp. Vol.* 67 (1993), pp. 213-33.

Boolos, G. (1997), "Is Hume's Principle Analytic?", in Heck (1997), pp. 245-62.

Boolos, G. (1998), *Logic, Logic and Logic* (Cambridge MA: Harvard University Press 1998).

Boolos, G. and Heck, R. (1998), "Die Grundlagen der Arithmetik §§ 82-3", in Schirn (1998), pp. 407-28.

Brandl, J. and Sullivan, P. (Eds.) (1998), *New Essays on the Philosophy of Michael Dummett* (Vienna: Rodopi 1998).

Burgess, J.P. (1984), Review of Wright (1983), *Philosophical Review* 93, 1984, pp. 638-40.
Clark, P. (1998), "Dummett's Argument for the Indefinite Extensibility of Set and Real Number" in Brandl and Sullivan (1998), pp. 51-63.
Demopoulos, W. (Ed.) (1995), *Frege's Philosophy of Mathematics*, Cambridge, MA: Harvard University Press (1995).
Demopoulos, W. (1998), 'The philosophical basis of our knowledge of number', *Noûs* 32 (1998), pp. 481-503.
Demopulos, W. (2000), On the origin and Status of Our Conception of Number (forthcoming).
Dummett, M. (1963), "The Philosophical Significance of Gödel's theorem", *Ratio* 5(1963), pp.140-55.
Dummett, M. (1991), *Frege: Philosophy of Mathematics* (London: Duckworth 1991).
Fine, K. (1998), "The Limits of Abstraction" in Schirn (1998), pp. 503-629.
Frege, G. (1879), *Begriffsschrift* (Halle: L. Nebert 1879).
Frege, G. (1884), *Die Grundlagen der Arithmetik* (Breslau: W. Koebner 1884); reprinted with English translation by J.L. Austin as *The Foundations of Arithmetic* (Oxford: Blackwell 1950).
Frege, G. (1893), *Die Grundgesetze der Arithmetik* vol.1 (Jena: H. Pohle 1893), part translated into English by Montgomery Furth in *The Basic Laws of Arithmetic* (Berkeley: University of California Press 1964).
Frege, G. (1902), Letter XV7[xxxvi/7] in Frege (1980), pp. 139-142.
Frege, G. (1980), Philosophical and Mathematical Correspondence, ed. G. Gabriel, 1980.
Hale, B. (1987), *Abstract Objects* (Oxford: Blackwell 1987).
Hale, B. (2000), "Reals by Abstraction", *Philosophia Mathematica* (2000).
Heck, R. Jr. (1993), "The Development of Arithmetic in Frege's *Grundgesetze der Arithmetik*", *Journal of Symbolic Logic* 58 (1993), pp. 579-601.
Heck, R. Jr. (1995), 'Frege's principle' in J. Hintikka *From Dedekind to Godel*, 1995, pp. 119-42.
Heck, R. Jr. (1997a), *Language, Thought and Logic* (Oxford: Oxford University Press 1997).
Heck, R. Jr. (1997b), "The Julius Caesar Objection", in Heck (1997a), pp. 273-308 (a).
Heck, R. Jr. (1997c), "Finitude and Hume's Principle", *Journal of Philosophical Logic* 26 (1997), pp. 589-617 (b).
Hodes, H. (1984), "Logicism and the Ontological Commitments of Arithmetic" *Journal of Philosophy* 81 (1984), pp. 123-49.
Parsons, C. (1964), "Frege's Theory of Number" in *Mathematics to Philosophy*, pp. 150-75.
Potter, M. (2000), *Reason's Nearest Kin*, (Oxford: Oxford University Press 2000).
Schirn, M. (1998), *Philosophy of Mathematics Today* (Oxford: Clarendon Press 1998).
Shapiro and Weir (1999), "New V, ZF and Abstraction", (2000), *Philosophia Mathematica* 1999, pp. 293-321.
Shapiro and Weir (2000), 'Neo-logicist Logic is not epistemically Innocent', *Philosophia Mathematica* 2000, pp. 160-189.
Smiley (1981), Frege and Russell, *Epistemologica*, 1981, pp. 51-56.
Wright, C. (1983), *Frege's Conception of Numbers as Objects*. (Aberdeen: Aberdeen University Press 1983).
Wright, C. (1997), "The Philosophical Significance of Frege's Theorem" in Heck (1997), pp. 201-45.
Wright, C. (1998a), "On the Harmless Impredicativity of N= (Hume's Principle)" in Schirn (1998), pp. 339-68.
Wright, C. (1998b), "Response to Dummett" in Schirn (1998), pp. 389-406.
Wright, C. and Hale, B. (2000), *The Reasons's Proper Study* (Oxford: Oxford University Press 2001).

Philosophy Department
University of St. Andrews
St. Andrews, Fife, Scotland
KY169AL
UK
pjc@st-and.ac.uk

MICHEL PATY

REMARKS ABOUT A "GENERAL SCIENCE OF REASONING"

Comments on Peter Clark's "Frege, Neo-Logicism and Applied Mathematics"

1.

As I am not at all a specialist on Gottlob Frege's work, my comments intended initially to be focused on an aspect that emerges in the last part of Peter Clark's paper "Frege, neo-logicism and applied mathematics"[1], where he treats the question of "applied mathematics" – an aspect that appealed to me and that was triggered by Frege's relationship between numbers and concepts, and reasoning. Starting with this concern, I have been led by my subject to propose some considerations about foundations and rationality which will go – briefly – in two directions. The first direction is that of a distinction between *logical* and *rational* foundations, whilst the second direction is that of taking into account as a fact *the historical development* of mathematics among the sciences, which modifies the terms of any foundational program. In delineating these considerations, I found that what I had in mind could apply to mathematics itself as well as to "applied mathematics", thus deviating somewhat from my first explicit intention. In conclusion I shall consider the possibility of a rational foundation programme for mathematical and physico-mathematical sciences which would take into account the changes in the scientific contents and the widenings of the forms of rationality that, in my view, make these changes possible. Such foundations for knowledge would not be any more static, but dynamical and would be possibly considered only retrospectively: they would be "forward foundations", in a sense that will be discussed in detail elsewhere[2].

2.

Let us first start from Frege's concerns with his logicist programme for mathematics and for science[3]. According to him, the logical foundation of arithmetic (if it were possible, as he hoped) would entail that *arithmetic* can be seen as akin to a *general science of reasoning about objects*. This is made evident by Peter Clark's statement: "To the question 'Why does arithmetic apply to reality?', the logician provides the clear answer: 'because it applies to everything that can be

F. Stadler (ed.), Induction and Deduction in the Sciences, 185–193.
© 2004 *Kluwer Academic Publishers.*

thought'. It is the most general science possible." And Peter Clark emphasizes: "In the simplest case for which the question arises – the application of cardinal numbers – [Frege's] solution is that arithmetic is applicable to reality because the concepts, under which things fall, themselves fall under numerical concepts."[4]

A question arises at this stage: is not this strong connection – a quasi identification – between concept and number, related to the fact that Frege's conception of knowledge is, as far as I know, a pre-Kantian realism, according to which objects that are described by thought correspond directly to objects that exist in reality? And objects that exist can be counted, this being their first qualification. Thought would begin by counting, and it would be possible for thought to construct everything from numbers.

To this rather simple view, one can object that the entities that are described by scientific thought (mathematics, physics, etc.) are never known directly, but through our mental and conceptual elaborations: the concepts by which we represent them are not only numerical ones. They are symbolic representations, and not all symbols are numbers. Relations of symbols can make sense without being relations of numbers (such are, for example, letters of the alphabet clustered in words).[5] Numbers, as such, are too restricted to be able to represent any kind of possible objects. Even if we restrict ourselves to mathematical – and "applied mathematical" – knowledge we see that other types of mathematical entities can be invoked to express concepts which are of the quantity type and of the continuous quantity type, as for example geometric or algebraic, including differential ones for continuity and change. Moreover, these quantities are not only related to "measure" (in the meaning of the word in 17th century, i.e. proportions, ratios), but also to respective positions (as Descartes said, "order", in his 1627 *Rules for the direction of the mind*); or, as Riemann wrote in his 1854 dissertation on the foundations of geometry, speaking on the properties of manifolds, they are not concerned only with metrics, but also with topology (*analysis situs*).

As to the applicability of mathematics, it is directly connected with the various possible types of magnitudes or quantities which can be used in the "application" field (for instance, physics). Take, as an example, geometrical reasoning, and, in particular, the qualitative study of curves represented by systems of differential equations as inaugurated by Poincaré at the end of 19th century. Exact solutions, with numbers, are overrun by qualitative ones, which reasoning can grasp much more powerfully. This shows how mathematical reasoning (and consequently, reasoning in other scientific disciplines where mathematical thought is used) is not restricted to numbers. To these considerations, one can add quantum theory, be it quantum mechanics or quantum field theories, and the mathematics that goes with it, i.e., state vectors of Hilbert space and linear operators acting on them with their non-commuting properties. It is difficult to reduce such conceptual entities merely to numbers. They actually are considered in their complex mathematical form, which entails their relational capacities, by mathematicians or theoretical physicists (and even by experimental physicists[6]) when they think about them and operate with them. And the common remark

that "everything ends in numbers" if we evoke, for instance, numerical approximations for the solutions of equations, or for the comparisons with experimental data, or computer binary calculations for simulating problems, is not a satisfactory answer to the question of the intelligibility of such mathematical or mathematized entities, because it is a practical and not a principle and foundational one.

The impossibility of grounding mathematics, and even arithmetic, on pure logic, is at least coherent with such a state of things whose consciousness comes to us from the lessons of mathematical and physico-mathematical reasoning practice.

3.

After this preliminary comment let us come to the *logic-versus-reason* consideration. The basic question about Frege's program seems to me to be the following: is reason (generally speaking, as the mental function for understanding) to be fundamentally identified with logic? (and consequently, arithmetic?) As we know, the answer is: no; because already with mathematics we cannot identify mathematical reasoning with arithmetic alone (or refer to it alone) as we have argued, and even less with logic alone, as Gödel's theorem demonstrates.

The question can also be considered from the point of view of the intellectual activity in the mathematical field for which many mathematicians see a kind of duality and competition between *logic* and *intuition*, the first corresponding to the requirement of *rigor*, the second to the process (and necessity) of *invention*. Mathematical theories are theories about some mathematical content, and this mathematical content is not given by purely logical operations of thought. It implies some *intuition*, for its invention as well as for its understanding: *intuition* is a word – indeed, a concept – that many mathematicians and theoretical physicists, and also philosophers, have used since the beginnings of philosophy, with a meaning that is not fixed from one to another and through the times. This is one of the reasons why the logicians would like to eliminate it from the language of metamathematics, and most contemporary philosophers from the language of the philosophy of knowledge. But its permanent use through the ages – even with varied meanings, from Aristotle to Descartes, Kant, Poincaré, Hilbert, Weyl, Einstein –, testifies to the necessity of having a philosophical concept to maintain an aspect of knowledge, and of even the most precise and exact knowledge, which escapes mere reduction to logic.

Poincaré's statements about the opposition between the two mathematical dispositions of logic and intuition are well known: "Logic is for demonstration, intuition for invention"[7]. To Poincaré, a geometer who would lack intuition would be in the same situation as a grammarian who, although he would know all of grammar, would lack ideas. Hilbert himself expressed some similar ideas, which appears interesting and significant as coming from the pioneer of the

axiomatization of geometry. He stated, in 1922, with a reference to Kant, that the sound matter of mathematics is given independently of any logic and that consequently mathematics will never be founded by logic alone, evidencing henceforth the reason for the failure of Frege and Dedekind (this being stated ten years before Gödel's theorem). To him, Hilbert, on the contrary, "the beforehand condition for applying logical *reasonings* is the presence of something given in the representation, some extralogical objects which intuitively happen to be there as an immediate experience, previous to any thought. (...) If the logical deduction is to be made firm, it has to bear on objects which can be grasped on the basis of all their sides, and which are such that their distinctive signs, their reciprocal relationships be identifiable with them, as something which is not reducible to anything else and which does not need it."[8]

The non-reducibility of the *mathematical content* (i.e., what is given in the mathematical relationships) to pure logic entails, in my view, that intuition in the sense of Poincaré and Hilbert pertains to reasoning: in no case is it erased under psychological considerations. For these mathematicians (as for the physicist Einstein), intuition is a kind of synthetic apprehension of the given by the understanding, this apprehension fully pertaining to what we call reasoning, which refers to "rationality".

In a way, that one has to hold on both *logic* and *intuition* in mathematical reasoning, despite the opposition of these two functions of the mind, indicates that a structured content of knowledge cannot be rendered by its structure alone. The theoretical and conceptual content of mathematics (or of any rational knowledge) cannot be generated by its logical structuration. Actually, it is something more of the contrary: although we can recognize *a posteriori* that the structure is co-generated with the content, in our knowledge the structure comes posterior to the mathematical content. To us, the logical structure emerges from the content once the latter is produced and made known to us: the (mathematical) content shows itself endowed with a logical structuration. When dealing with knowledge, we must consider the *effective science* which is produced and made intelligible, and when considering it, one cannot escape the fact of mathematical or scientific invention or creation. Even if mathematical reality (as all mathematical contents taken together) or physical reality (that one to which physics is referred, in the "real, external, world") are considered separately from human mind and objectified, as knowledge they are the product of intellectual mind activity. Any "foundational program", and the logical one itself, is a program about a given knowledge and cannot ignore these dimensions of knowledge. The invention process from which the (mathematical or else) knowledge content has been issued precedes the recognition of its logical structuration. One cannot invent, when reasoning, with logic alone, even if logic is implicitly present in this reasoning. For reasoning operates with objects, statements, symbols that are not univoquely given, neither at the start nor in the course of its operation. The mind (its function of understanding) has to choose and eventually to connect between them entities (propositions) that were previously unconnected.

The mathematical philosopher Jean Cavaillès[9] stated something to this effect in his 1938 doctoral dissertation, *Méthode axiomatique et formalisme. Essai sur le problème du fondement des mathématiques*: "Mathematics is richer than logic, insofar as it is an effective thought and any effective thought supposes applying abstract thought to an *intuition*"[10]. "To reject or to found a theory", he wrote in his *Remarques sur la formation de la théorie abstraite des ensembles* (his complementary doctoral dissertation), "is neither definitive nor devoid of degrees; (...) it cannot be done simply by logical investigation; (...) the pragmatic considerations of the militant mathematician tell the last word"[11]. As for him, according to another mathematical philosopher, by the way a Cavaillès disciple, Jean-Toussaint Desanti, "he wanted to catch the "mathematical experience", in its necessary movement of constitution, and in it, consubstantial to this necessity, logic itself"[12]. And Gilles-G. Granger (also a former Cavaillès disciple) emphasizes that for Cavaillès, the historical development of mathematics is rational and there is an objectivity of the mathematical becoming, while noting the fascinating paradox of the history of mathematics, which is at the same time imprevisible and rational.[13] Cavaillès himself, who had a Spinozian conception of an immanent development of mathematics, stated in his posthumous book *Sur la logique et la théorie de la science*, that "the true meaning of a theory does not stand in an aspect that is understood by the scientist himself as essentially provisional, but in a conceptual becoming which cannot be stopped"[14].

4.

We could transform Frege's worry about identifying reasoning with dealing-with-numbers into the following one: does reasoning in general have a direct connection with mathematical reasoning? We consider the question, this time, taking mathematics as any kind (and all the kinds) of mathematical theories and concepts, those which are known to us, but also taking into account the fact, taught to us by the effectiveness of mathematics, as Cavaillès recalled, that mathematics are not closed and that more mathematics are to be invented. We are tempted to say, pursuing the parallel: more mathematics through more reasoning, more rationality modes, which are to be invented with them. For reasoning, and rationality, considered in general is (possibly) akin to mathematical reasoning, which includes the extension (by the creativity of the mind) of mathematics (inasmuch as rationality, in its exercising, implies the extension of what we are used to call rationality, or modes of rationality).

It is clear, from all what precedes, and from many other considerations which we cannot exhaust in this short space, that reasoning, either in mathematics or generally speaking, is definitely not to be identified with logic because logic is too restrictive and cannot by its only exercise, when one is reasoning about something, modify or generate the premises, for example when building theories (mathematical or physical ones, or other kinds of theories).

That reasoning is not reduced to logic, or even not based on logic, does not mean, clearly, that elaborating scientific theories would be illogical for that reason. Logic is still present, be it implicitly, as a kind of regulation, a criterion along the path of the "working thought" [15] which makes emerge knowledge and forms of knowledge. Actually, we are no longer, with this kind of considerations, in the domain of *foundations* of knowledge in the usual sense that considers foundations that would be already given, but in the domain of the extension (or growing) of knowledge. A question arises whether one can speak again of "foundations" when one knows that sound knowledge is subject to modifications, and even more, that it would not be knowledge in the full sense if it were to stay static. Can we consider with some meaning a *foundation* for a knowledge which has, as a fundamental property, to be *dynamical*, i.e., to be modified with time? If this could be considered, it would have to be on the condition of thinking anew the concept of *foundation*. The property of stability is generally required for foundations, and we know that mathematical and more generally scientific knowledges are relatively stable, inasmuch as they are growing and in a way cumulating. New knowledge increases the precedent one, and its discovery and formulation base themselves on the previous one, whilst at the same time it can entail a complete change of meaning of concepts, theories and bodies of knowledge.

In short, stability has to cope with instability. Is this foundation? We shall not solve the problem here, but only note that the exercise of reason is in all cases the only instance of judgement: reason and its operating forms of rationality, mathematical, physical, etc., and common ones as well. These forms themselves are changing, actually they are enlarging, as we suggested, to make the new knowledge possible and make it intelligible, and to continue understanding the previous ones, but in a somewhat modified way, from another point of view (generally, a more unifying one). The forms of rationality are changing, but their function stands and, with the function, something stable through its own motion and modifications. Such a problem can well be considered as a problem of foundation, not for logic but for reason. In renewed Kantian terms: what are the conditions in the rational structures of the mind for a dynamical and at the same time relatively secure knowledge be possible?

5.

To complete what has been sketched above, let us consider the question of the "application of mathematics" in its relation to the foundational problem and with the problem of the generalization of aspects of mathematical reason to reason in general. To say something of it in a few words, the so-called "application of mathematics", let us say in physics, is determined by the external world which manifests itself empirically. But, at the same time, this "application" (or to say it better, this "use of a mathematical tool for reasoning") transforms the empirical,

by assimilating it, into rational constructions and representations[16]. The legitimation of this transformation of an hypothetical-empirical into a rational-intelligible is given by its success in representing the physical phenomena, but also in anticipating them, and a condition for it, from the theoretical point of view, is the systematical character, which requires that the representation constitutes a *theoretical system.* (Examples are well-known: classical mechanics, relativistic electrodynamics, general relativity, thermodynamics, and we can add quantum physics[17]).

The justification of this process of rationalization does not therefore come from *beneath* (from static foundations), neither does it come from *before* (as if all present and future knowledge was already contained in its previous forms), but it comes from *forward.* I mean that we know the justification and foundation (as firm ground) of a given piece of theoretical knowledge only when it has been already achieved or completed. Maybe this situation is also that one of non-empirical knowledge such as mathematics, as its seem to appear from what we have said previously. It seems indeed rather reasonable to think that if knowledge is taken as a dynamical process, and is unachieved, unclosed, its true foundation has not been already attained, and is not yet attainable with the means that we presently dispose of.

A last remark which would need larger developments, about formal considerations linked to the sciences of nature, which may clarify a legitimate use of the expression "rational reconstruction". This expression, used for instance by Imre Lakatos, Elie Zahar and others, has the inconvenience of leading one to think that construction of science by itself is not rational (indeed, Reichenbach, Popper and many others thought so) for it needs a "rational reconstruction", when we have suggested, on the contrary, that even intuition in creative scientific work sits on the side of the rational. From what has been said about "mathematical application" on other sciences – and essentially physics is concerned –, we can then understand the meaning of a kind of "rational reconstruction" in the same way as we understand the meaning of "axiomatizing a theory". Both are meaningful, not because there would have been a lack of rationality in the creation and in the discovery of new knowledge, or in a non-axiomatized theory. (Rational) reconstruction and axiomatization actually consider the given theory from another point of view, the point of view of the "economy" of the propositions and of the theoretical structure. In these, the theoretical structure is rearranged in such a way as to show a deductive and logical sequence from transformed premises or starting principles or axioms. These premises, principles, axioms, carry the physical content (in the case of this science) which can be evaluated, and can eventually be considered henceforth as the new basic (foundational) and "natural" (or "reasonable") concepts. The correlative achievement of a direct rational-and-logical view can be seen, as for itself, as a change – indeed, an extension – of the forms of rationality operating on this body of knowledge.

The sense of these reflections is that, despite the failure of the foundational program in the logical sense, there remains a meaning for a "foundational" concern, which is grounded in the function of rationality, that makes mathematical and scientific knowledge not only something known to us, but deeply intelligible to us.

NOTES

1. Peter Clark, "Frege, neo-logicism and applied mathematics", in this volume.
2. Michel Paty, "Des fondements vers l'avant. Sur la rationalité des mathématiques et des sciences formalisées", Contribution to the *Colloque International "Aperçus philosophiques en logique et en mathématiques. Histoire et actualité des théories sémantiques et syntaxiques alternatives"*, Nancy, 30 Sept.-4 Oct. 2002 (forthcoming).
3. Gottlob Frege, *Die Grundlagen der Arithmetik. Eine logisch-mathematische Untersuchung über den Begriff der Zahl*, Breslau, W. Koebner, 1884 (Engl. transl. by J.L. Austin, *The Foundations of Arithmetic*, Oxford, Blackwell, 1950. French transl. by C. Imbert, *Les Fondements de l'arithmétique*, Paris, Seuil, 1970); Gottlob Frege, *Ecrits logiques et philosophiques*, transl. in French from German, Paris, Seuil, 1971; J. van Heijenoort, *From Frege to Gödel. A source Book in Mathematical Logic (1879-1931)*, Cambridge (Mass.), Harvard University Press, 1967.
4. Peter Clark, *ibid.*
5. It is true that Frege himself considered a more general symbolic notation than pure numbers (ideography).
6. Michel Paty, "The concept of quantum state: new views on old phenomena", *in* A. Ashtekar, R.S. Cohen, D. Howard, J. Renn, S. Sarkar & A. Shimony (eds.), *Revisiting the Foundations of Relativistic Physics: Festschrift in Honor of John Stachel*, Boston Studies in the Philosophy and History of Science, Dordrecht, Kluwer, forthcoming.
7. Henri Poincaré, "La logique et l'intuition dans la science mathématique et dans l'enseignement", *L'Enseignement mathématique*, 1, 1889, 157-162 (Repr. in Henri Poincaré, *Œuvres*, Paris, Gauthier-Villars (11 vols., 1913-1965), vol. 11, pp. 129-133.
8. David Hilbert, "Neubegründung der Mathematik" (1922), in David Hilbert, *Gesammelte Abhandlungen*, Berlin, B. 3, 1935.
9. Jean Cavaillès (1903-1944), prematurely carried off to death by the Nazis in occupied France, where he was involved in Resistance activities, was a philosopher and logician.
10. Jean Cavaillès *Méthode axiomatique et formalisme, Essai sur le problème du fondement des mathématiques* (Thesis, 1937, 1rst ed, 1938), Introduction by Jean-Toussaint Desanti, Preface by Henri Cartan, Paris, Hermann, 1981. (My emphasis, MP).
11. Jean Cavaillès, *Remarques sur la formation de la théorie abstraite des ensembles* (Complementary Thesis, 1937, 1rst ed, 1938), *in* Jean Cavaillès, *Philosophie mathématique*, Preface by Raymond Aron, Introduction by Roger Martin, Paris, Hermann, 1962, pp. 23-174.
12. Jean-Toussaint Desanti, "Souvenir de Jean Cavaillès", *in* Jean Cavaillès, *Méthode axiomatique et formalisme. Essai sur le problème du fondement des mathématiques* (1981 ed.). Introduction by Jean-Toussaint Desanti, Preface by Henri Cartan, Paris, Hermann, 1981.
13. Gilles-Gaston Granger, *Science, langage, philosophie*, Collection "Penser avec les sciences", EDP-Sciences, Paris, 2003, chapter on "Jean Cavaillès et l'histoire", pp. 76-84.
14. Jean Cavaillès, *Sur la logique et la théorie de la science* (written in 1942, 1rst ed., 1946), 3rd ed., Paris, Vrin, 1976.
15. By the expression "working thought" I mean what I would call in French: *la pensée au travail*. The idea of scientific or rational thought as a working action has been developed by Gilles-Gaston Granger, notably in his *Essai d'une philosophie du style*, Paris, Armand Colin, 1968; re-ed., Paris, Odile Jacob, 1988.

16. Michel Paty, "Intelligibilité et historicité (Science, rationalité, histoire)", *in* J.J. Saldaña (ed.), *Science and Cultural Diversity. Filling a Gap in the History of Science*, Cadernos de Quipu 5, México, 2001, pp. 59-95; "Les concepts de la physique: contenus rationnels et constructions dans l'histoire", *Principia* (Florianopolis, Br), 5, n°1-2, junho-dezembro 2001, 209-240 (English version: "The concepts of physics: rational contents and constructions in history", in J. Margolis and T. Rockmore (eds.), forthcoming.
17. On the last one, see in particular: Michel Paty, "La physique quantique ou l'entraînement de la forme mathématique sur la pensée physique", *in* C. Mataix y A. Rivadulla (eds.), *Física cuantica y realidad. Quantum physics and reality*, Madrid, Editorial Complutense, 2002, pp. 97-134; " The concept of quantum state: new views on old phenomena", *op. cit.*

Equipe REHSEIS, CNRS et Université Paris 7 – D. Diderot
Centre Javelot, 2 Place Jussieu
F 75251 Paris – Cedex 05
France
paty@paris7.jussieu.fr

JEAN-JACQUES SZCZECINIARZ

TWO QUESTIONS ABOUT THE REVIVAL OF FREGE'S PROGRAMME

Comments on Peter Clark's "Frege, Neo-Logicism and Applied Mathematics"

INTRODUCTION

The essential aim of Professor P. Clark's paper is to show that good reasons exist for opening the way for a revival of Frege's Programme. Frege's Programme tend to show that numbers are given to us by explaining the meanings of identity statements in which numbers words occur.

The most powerful tool is Second Order Logic together with "Basic Law Five". It is a very interesting fact that (whether or not it was stated by Frege as theorem) "the Dedekind-Peano axioms postulates for number theory in their second order form can be derived from a combination of Second Order Logic and Hume's Principles." The arguments show that some worry remains in all the proposed issues to dissolve the difficulties met with the fullfilment of the Neo-Fregean Programme.

1.

The question concerns the status of Second Order Logic particularly with respect to the claim that Hume's Principle and other abstraction principles can be thought of analytic stipulations introducing the numbers names. For instance, what guarantees, Professor Clark asks, the correct reading of the stipulation to introduce the name "zero" to designate the unique object,

$$(\exists!\, y(\mathrm{N}x{:}\neg(x = x) = y)$$

2.

A question also arises at to whether it is only natural or not that one asks "is there such a number as anti-zero"? Second Order Logic allows us to prove that the successor relation is functional and one-to-one and that we are able to prove that every number has a successor.

F. Stadler (ed.), Induction and Deduction in the Sciences, 195–197.
© 2004 *Kluwer Academic Publishers.*

We cannot show within the theory (According Frege's Arithmetic) that *Finite (n) $\neq$ *(x=x) (* is a total function relating concepts to objects).

The worry about an anti-zero can be resolved. The function * is not, Professor Clark says, and Wright as well, to be regarded as having as its domain all concepts. It is to be confined to sortal concepts (as in Frege's prime tenet).

As Wright, Professor Clark claims:

So self-identity is not a sortal concept. If we take it that, save where F is assured an empty extension on purely logical grounds, only sortal concepts, and concepts formed restricting a mere predicable to a sortal concept, have cardinal numbers, it follows that there is no universal number.

3.

The most important difficulty looms. We cannot consider the sortal concepts "set", "ordinal", "cardinal" according to one of Hume's Principle as having a number associated with them, because according to ZFC there are no numbers associated with the concepts "set", "ordinal", "cardinal number". Boolos remarks:

The worry is that the theory of numbers we have been considering is incompatible with ZF set theory plus standard definitions, on the usual and natural readings of the non-logical expressions of the both theories ...

The first conception states that every sortal concept has a number and, according to the other, says that every sortal concept has a number if and only if the extension of the concept is a set.

However far one goes, one will not find means to conciliate both concepts of the cardinal number. This situation would amount to giving up the foundational programme. I carefully suggest that this impenetrable problem, Professor Clark says, ("there are two conceptions of cardinal number one for concepts and one for sets" is the eclectic version), may push us to evaluate other making of number, particularly the ones trying to obtain structures or quasi structures inside of set theoretic number. [Conway, Gonshor 86]

4.

Wright writes that:

the Frege's Programme should have a determinate cardinal number just (anti-Fregean) when they compose a set.

And he proposes that the concepts should be restricted to those which are 'definite', i.e., 'not indefinitely extensible'.

It is very interesting that Dummett has recently revived this idea, which is the famous Russell's Vicious Circle Principle:

All our contradictions have in common the assumption of a totality such that, if it were legitimate, it would at once be enlarged by new members defined in terms of itself.

The Russell of this period agreed that the objects falling under such self-reproductive properties like ordinal or set form no totality.

I would like to question about this revival: Russell later introduced the axiom of reducibility, according to which to every function of a higher order there exists a corresponding function of the first order which is extensionally equivalent to it. He sometimes defended this axiom as being of the same kind as Zermelo's axiom of choice or Indefinite extensibility. I would like to know whether it is possible or not to get a revival of this idea without following the same logical trajectory?

The very interesting conception of Wright about a New Hume leads to a new form of dilemma. It would be interesting in my view, although still asking to formalise the fact that we are necessary prisoner of the set'notion (or prejudice?). I can only stand here with the intuitive aspect of my question. It was my first question.

My second question is about the neologicist trend and the revival of Frege's Programme. I wonder whether they are nearer the arithmetic trend for which the number's concept is in a certain sense a diagrammatic and structural concept. The deepest mathematician who snatched the number from the logicist hands was Kronecker [Kronecker 1886].

On the other side, the history of the Category Theory (Cartan, Eilenberg [56] can be understood if we take into account Russell's paradox and consequently the falsity in a certain sense of the Law Five. One may ask if we can change this situation inside of the Neo-Fregean Programme?

ADDITIONAL BIBLIOGRAPHY

[Cartan Eilenberg 56] Henri Cartan, Serge Eilenberg *Homological Algebra*, Princeton University Press, Princeton 1956.

[Gonshor 86] Harry Gonshor *An Introduction to the Theory of the Surreal Numbers,* Cambridge University Press 1986.

[Kronecker 1886] Leopold Krönecker *Über den Zahlbegriff* in *Mathematische Werke,* t. III pp. 251-274.

Université de Bordeaux 3 – Michel de Montaigne
et Equipe REHSEIS, CNRS Université Paris 7 – Denis Diderot
2 Place Jussieu
75251 Paris
France
szczeciniarz@paris7. jussieu.fr
jjs@mathp7.jussieu.fr

DONALD GILLIES

HANDLING UNCERTAINTY IN ARTIFICIAL INTELLIGENCE, AND THE BAYESIAN CONTROVERSY

1. INTRODUCTION

This paper is divided into two parts. In the first part (sections 2 and 3), I will describe briefly how advances in artificial intelligence (AI) in the 1970s led to the crucial problem of handling uncertainty, and how attempts to solve this problem led in turn to the emergence of the new theory of Bayesian networks. I will try to focus in this historical account on the key ideas and will not give a full account of the technical details. Then, in the second part (section 4), I will consider the implications of these new results for the long-standing controversy between Bayesians and non-Bayesians.

2. THE BREAKTHROUGH WITH EXPERT SYSTEMS IN THE 1970S

Research in AI began in the 1950s and many important ideas were developed by the pioneers. Then in the 1970s, a breakthrough was produced by the creation of expert systems. The lead here was taken by the Stanford heuristic programming group, particularly Buchanan, Feigenbaum, and Shortliffe. What they discovered was that the key to success was to extract from an expert the knowledge he or she used to carry out a specialised task, and then code this knowledge into the computer. In this way they were able to produce 'expert systems' which performed specific tasks at the level of human experts. One of the most important of these early expert systems (MYCIN) was concerned with the diagnosis of blood infections. This system will now be briefly described, and it will then be shown that its implementation led to the problem of how to handle uncertainty in AI.

MYCIN was developed in the 1970s by Edward Shortliffe and his colleagues in collaboration with the infectious diseases group at the Stanford medical school. The medical knowledge in the area was codified into rules of the form: IF such and such symptoms are observed, THEN likely conclusion is such and such. MYCIN's knowledge base comprised over 400 such rules which were obtained from medical experts. An example of such a rule will be given in a moment, but first it would be as well to present some evidence of MYCIN's success.

F. Stadler (ed.), Induction and Deduction in the Sciences, 199–216.
© 2004 *Kluwer Academic Publishers.*

To test MYCIN's effectiveness, a comparison was made in 1979 of its performance with that of nine human doctors. The program's final conclusions on ten real cases were compared with those of the human doctors, including the actual therapy administered. Eight other experts were then asked to rate the ten therapy recommendations and award a mark, without knowing which, if any, came from a computer. They were requested to give 1 for a therapy which they regarded as acceptable and 0 for an unacceptable therapy. Since there were eight experts and ten cases, the maximum possible mark was 80. The results were as follows:[1]

MYCIN	52	Actual therapy	46
Faculty-1	50	Faculty-4	44
Faculty-2	48	Resident	36
Inf dis fellow	48	Faculty-5	34
Faculty-3	46	Student	24

So MYCIN came first in the exam, although the difference between it and the top human experts was not significant.

Let us now examine one of MYCIN's rules. The following rule is given by Shortliffe & Buchanan:[2]

"If: (1) the stain of the organism is gram positive (S_1), and
(2) the morphology of the organism is coccus (S_2), and
(3) the growth conformation of the organism is chains (S_3)
Then: there is suggestive evidence (0.7) that the identity of the organism is streptococcus (H_1)"

In symbols, this could be written: If S_1 & S_2 & S_3 , then there is suggestive evidence p that H_1, where p = 0.7. Here, S_1, S_2, S_3 are the observations / symptoms, which support hypothesis H_1 to a particular degree. These rules were obtained from the medical experts. The numbers they contain, such as 0.7, look like probabilities, and they too were obtained from the experts. The expert was in fact asked: "On a scale of 1 to 10, how much certainty do you affix to this conclusion?" The answer was then divided by 10.

It looks as if Shortliffe & Buchanan are using probability in the subjective sense to measure the degree of personal belief held by an expert. This at once raises the question of why subjective probabilities obtained from experts are preferred to objective probabilities obtained from data. Shortliffe & Buchanan do consider this question, and they answer[3] that in typical medical applications there is not enough data to obtain the requisite objective probabilities. This in turn is because of the inadequacy of hospital records, and the changes which are continually occurring in disease categories. It is interesting to note that only three years previously, another group working on computer diagnosis had reached

exactly the opposite conclusion. This research group, working in Leeds, was headed by de Dombal. Their results are contained in de Dombal *et al.* (1972)[4], and Leaper *et al.* (1972)[5]. I will consider them in the final section of the paper.

De Dombal's approach was, however, largely ignored for the next twenty years, and nearly every researcher in the field made use of subjective probabilities. There are two possible reasons for this. First of all, it may in many cases have been difficult to obtain objective probabilities from data. Secondly, the general methodology of expert systems research, since it involved obtaining knowledge from the experts, may have encouraged the idea of obtaining probabilities as the degrees of belief of these experts. I will discuss further the question of objective versus subjective probabilities in the final section of the paper. Let us now return to a consideration of MYCIN, and the sample rule given earlier.

So far we have rather assumed that the figure 0.7 in the rule from MYCIN is an ordinary probability, but this is not the case, as Shortliffe & Buchanan make clear in the following passage:[6]

> ... this rule at first seems to say $P(H_1 \mid S_1 \& S_2 \& S_3) = 0.7$, ... Questioning of the expert gradually reveals, however, that despite the apparent similarity to a statement regarding a conditional probability, the number 0.7 differs significantly from a probability. The expert may well agree that $P(H_1 \mid S_1 \& S_2 \& S_3) = 0.7$, but he becomes uneasy when he attempts to follow the logical conclusion that therefore $P(\text{not}.H_1 \mid S_1 \& S_2 \& S_3) = 0.3$. The three observations are evidence (to degree 0.7) *in favor* of the conclusion that the organism is a streptococcus and should not be construed as evidence (to degree 0.3) *against* streptococcus.

Shortliffe & Buchanan used this observation to motivate the introduction of a *non-probabilistic* model of evidential strength. Their measure of evidential strength was called a *certainty factor*, and certainty factors neither obeyed the standard axioms of probability theory, the Kolmogorov axioms, nor combined like probabilities.

Certainty factors were criticized by those who favoured a probabilistic approach, cf. Adams (1976)[7] and Heckerman (1986)[8], and in fact the next expert system we will consider (PROSPECTOR) did move more in the direction of standard probability.

PROSPECTOR, an expert system for mineral exploration, was developed in the second half of the 1970s at the Stanford Research Institute. A good general account of the system is given by Gaschnig in his 1982 article[9]. PROSPECTOR's most important innovation was to represent knowledge by an *inference network* (or *net*). This is motivated by Duda *et al.* in their 1976 as follows:[10]

> A collection of rules about some specific subject area invariably uses the same pieces of evidence to imply several different hypotheses. It also frequently happens that several alternative pieces of evidence imply the same hypothesis. Furthermore, there are often chains of evidences and hypotheses. For these reasons it is natural to represent a collection of rules as a graph structure or *inference net.*

A part of PROSPECTOR's inference network is shown in figure 1.

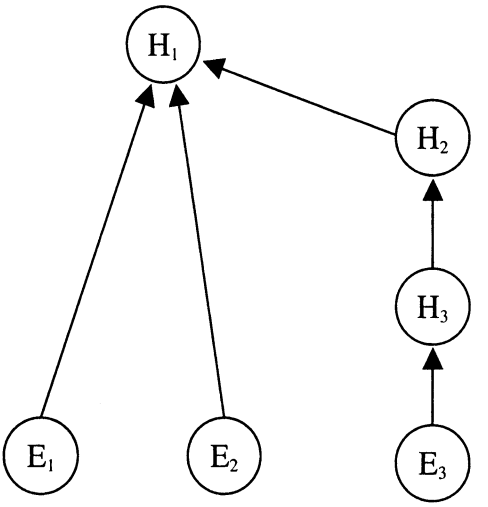


Figure 1

H_1 = There are massive sulfide deposits.
H_2 = There are clay minerals.
H_3 = There is a reduction process.
E_1 = Barite is overlying sulfide.
E_2 = Galena, sphalerite, or chalcopyrite fill cracks in rhyolite or dacite.
E_3 = There are bleached rocks.

Evidence E_1 is taken as supporting hypothesis H_1, and this is indicated by the arrow joining them in the inference network. Similarly, E_2 supports hypothesis H_1, while E_3 supports H_3 which supports H_2 which supports H_1. Note how these rather complicated relations are simply and elegantly represented by the arrows of the network. Each inference arrow has a strength associated with it, and this obtained from the expert as in the case of MYCIN.

PROSPECTOR, however, differs from MYCIN in using subjective Bayesianism rather than certainty factors. This subjective Bayesianism is not entirely pure, since, as in the case of MYCIN, it is combined with the use of fuzzy logic formulae. This use of fuzzy logic tended to disappear in further developments.

In PROSPECTOR, Bayesianism is formulated using odds rather than probabilities. The odds on a hypothesis H [O(H)] are defined as follows:

$$O(H) = P(H)/P(\neg H)$$

Writing down Bayes theorem first for H and then for ¬H, we get

$$P(H \mid E) = P(E \mid H)\, P(H) / P(E)$$
$$P(\neg H \mid E) = P(E \mid \neg H)\, P(\neg H) / P(E)$$

so dividing gives

$$O(H \mid E) = \lambda(E)\, O(H) \qquad (1)$$

where $\lambda(E)$ is the likelihood ratio $P(E \mid H)/P(E \mid \neg H)$. (1) is the odds and likelihood form of Bayes theorem, and it is used in PROSPECTOR to change the prior odds on H to the posterior odds given evidence E.

Let us now consider the problems which arise if we have several different pieces of evidence $E_1, E_2, \ldots, E_n$ say. We might in practice have to update using any subset of these pieces of evidence $E_i, E_j, \ldots, E_k$ say, where (i, j, ...k) is any subset of (1, 2, ... , n). If we use (1), this would involve having values of $\lambda(E_i \,\&\, E_j \,\&\, \ldots \,\&\, E_k)$ for all subsets of (1, 2, ...n). When we remember that, on this approach, the values of λ are obtained from the domain experts, we can see that obtaining the requisite values of λ is scarcely possible. Clearly, some simplifying assumptions are necessary to produce a workable system, and the designers of PROSPECTOR, therefore, made the following two conditional independence assumptions:

$$P(E_1, \ldots, E_n \mid H) = P(E_1 \mid H) \ldots P(E_n \mid H) \qquad (2)$$
$$P(E_1, \ldots, E_n \mid \neg H) = P(E_1 \mid \neg H) \ldots P(E_n \mid \neg H) \qquad (3)$$

Given these assumptions, the whole problem of updating with many pieces of evidence becomes simple, and, in fact,

$$O(H \mid E_1 \,\&\, \ldots \,\&\, E_n) = \lambda_1 \lambda_2 \ldots \lambda_n\, O(H) \text{ where } \lambda_i = \lambda(E_i)$$

The only remaining problem was whether the conditional independence assumptions (2) and (3) are plausible. The search for a justification of these assumptions led, as we shall see in the next section, to the modification of the concept of inference network, and the emergence of the concept of *Bayesian network*.

3. THE EMERGENCE OF BAYESIAN NETWORKS IN THE 1980S

The concept of Bayesian network was introduced and developed by Pearl in a series of papers: Pearl (1982[15], 1985a[17], 1985b[11], 1986[12]), Kim & Pearl (1983)[16], and a book: Pearl (1988)[18]. An important extension of the theory was carried out by Lauritzen & Spiegelhalter (1988)[19], while Neapolitan's 1990 book[31] gave a clear account of these new ideas and helped to promote the use of Bayesian

networks in the AI community. In what follows, I will comment on a few salient features of Bayesian networks which will be important when we consider their implications for the Bayesian controversy.

The actual term *Bayesian (or Bayes) network* was introduced in Pearl's 1985b where it is defined as follows:[11]

> Bayes Networks are directed acyclic graphs in which the nodes represent propositions (or variables), the arcs signify the existence of direct causal influences between the linked propositions, and the strengths of these influences are quantified by conditional probabilities.

This verbal account is illustrated by a diagram which is reproduced, with different lettering, in Figure 2.

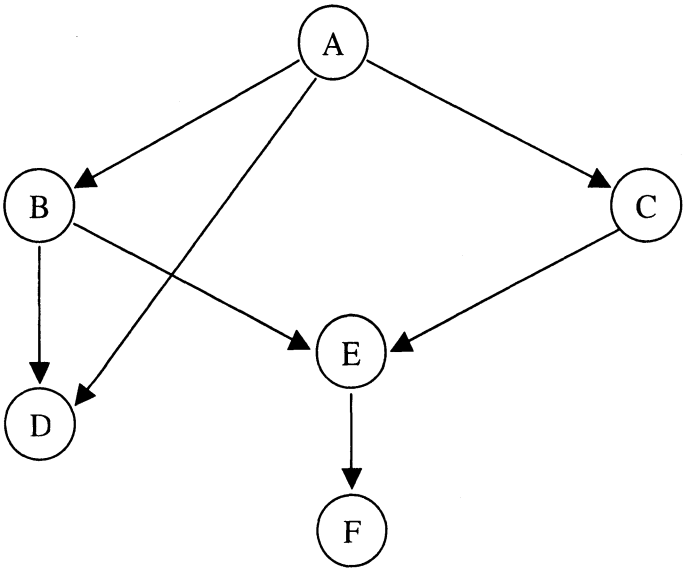


Figure 2

If we compare the network of figure 2 with that of figure 1, two differences should be noted immediately. First of all, the arrows in the inference network of figure 1 represent a relation of support holding between, e.g., E_3 and H_3, while the arrows in the Bayesian network of figure 2 represent causal influences, so that, e.g., the arrow joining A to B means that A causes B. Secondly, corresponding to the first difference, we can say that, in a certain sense, the arrows of a Bayesian network run in the opposite direction to those of an inference network. Pearl puts this point as follows:[12]

... in many expert systems (e.g., MYCIN), ... rules point from evidence to hypothesis (e.g., if symptom, then disease), thus denoting a flow of mental inference. By contrast, the arrows in Bayes' networks point from causes to effects or from conditions to consequence, thus denoting a flow of constraints in the physical world.

This reversal of arrows from inference networks to Bayesian networks is illustrated in Figure 3, which shows one pair of nodes taken from the portion of PROSPECTOR's inference network shown in Figure 1.

Reversal of Arrows

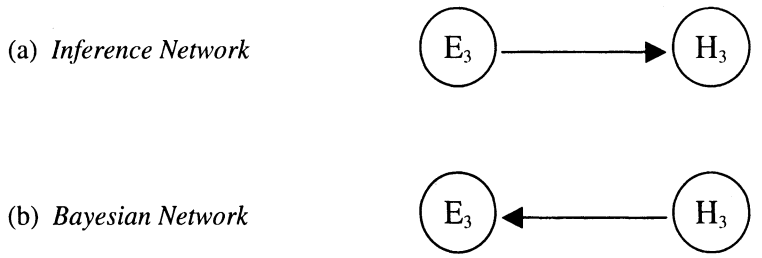


Figure 3

Here E_3 = There are bleached rocks, while H_3 = There is a reduction process. From the point of view of an inference network (a), we regard the evidence of bleached rocks as supporting the hypothesis that there is a reduction process, while, from the point of view of a Bayesian network (b), we regard there being a reduction process as a cause of there being bleached rocks. In his 1993[13], Pearl gives an account of his discovery of Bayesian networks, and says that one factor that led him to the idea was his consideration of the concept of influence diagrams introduced by Howard and Matheson (1984).[14] Pearl decided to limit the influences specifically to causal influences.

Let us now make a few further points about Bayesian networks. If, in such a network, an arrow runs from node A to node B, then A is said to be a parent of B, and B a child of A. If a node has no parents, it is called a root, so that in figure 2, A is a root. In a Bayesian network, it is possible for a child to have several parents. Thus in figure 2, E has parents B and C. If, however, every child has at most one parent, the network is called a tree. As in the earlier case of PROSPECTOR's inference networks, in order to make computation feasible, some conditional independence assumptions have to be made. For a Bayesian network, these are that a node is conditionally independent given its parents of the rest of the network except its descendants.

From his definition of Bayesian networks, Pearl developed algorithms which allow Bayesian updating to take place in such networks. If one of the variables which represents an observation is set to a particular value, the changes brought about by this new information in all the probabilities throughout the tree can be computed in an efficient manner. Pearl began in his 1982[15] by developing an updating algorithm for a simple form of network, namely a tree. He then extended his algorithm to more complicated networks. Kim and Pearl (1983)[16] generalised from trees to Bayesian networks which are singly connected, i.e., there exists only one (undirected) path between any pair of nodes. Pearl, in his 1986[12], tackled the further extension to Bayesian networks which are multiply connected. This problem was also investigated by Lauritzen & Spiegelhalter who, in their 1988[19], solved it using the idea of reducing a multiply connected network to a tree of cliques. Their algorithm has been generally adopted by the AI community.

Let us now turn from these powerful mathematical developments to the consideration of a conceptual point. It will have been noted that two rather different definitions of Bayesian network have been given. The first definition is in terms of causes. Thus, in figure 2, the arrows are taken as denoting a causal link between the two nodes which they join. The second definition is by contrast purely probabilistic. In figure 2, the variables A, B, C, D, E, F are taken to be random variables with a joint probability distribution, and the network becomes a Bayesian network if the relevant conditional independence assumptions are satisfied. I will henceforth use the term 'Bayesian network' for networks defined purely probabilistically in the manner just explained, and call the networks defined in terms of causes: 'causal networks'. Pearl tends, however, to use the terms 'Bayesian network' and 'causal network' interchangeably, because he believes the two notions to be closely connected. More specifically, his idea is that if in a network, the parents of every node represent the direct causes of that node, then the relevant conditional independence assumptions will automatically be satisfied. As he says:[20]

> Causal utterances such as 'X is a direct cause of Y' were given a probabilistic interpretation as distinctive patterns of conditional independence relationships that can be verified empirically.

A suggested link between causality and conditional independence in fact goes back to Reichenbach (1956).[21] Reichenbach considers two events, B and C, which are correlated. For example, in a travelling troupe of actors, B = the leading lady has a stomach upset, and C = the leading man has a stomach upset. We can explain such correlations, according to Reichenbach, by finding a common cause, namely that the leading lady and the leading man always have dinner together. The common stomach upsets occur when the food in the local restaurant has gone off. Denote 'dining together' by A. We then have the causal graph shown in figure 4.

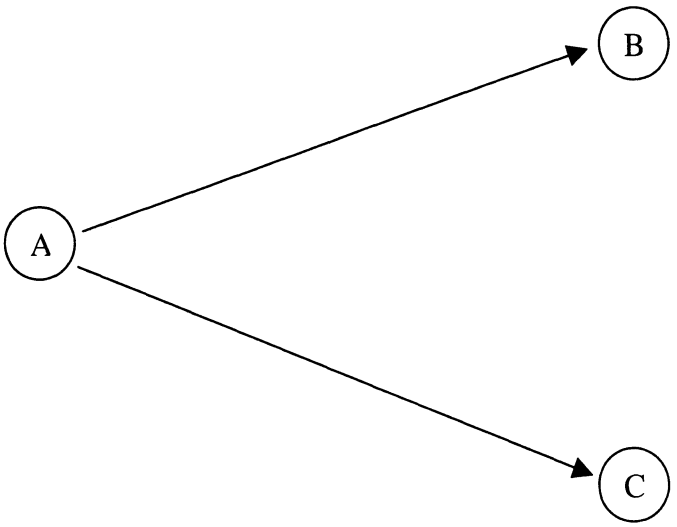


Figure 4

Reichenbach then claimed that, conditional on A, B and C were no longer correlated but independent, i.e., P(B&C | A) = P(B | A) P(C | A). He also expressed this idea by saying that a common cause A screens one of its effects B off from the other C. Reichenbach's causal fork is just a simple case of a Bayesian network. We can indeed apply his term 'screening off' to Bayesian networks by saying that in such networks, the parents of a node screen it off from all the other nodes in the network except its descendants.

We are now in a position to summarise the ingenious way in which Bayesian networks solved the problem of handling uncertainty in expert systems. In most of the domains considered, e.g., medical diagnosis, a domain expert is very familiar with the various causal factors operating. It should, therefore, be an easy matter to get him or her to provide a causal network. By the addition of probabilities, this can be turned into a Bayesian network. In earlier systems, such as MYCIN or PROSPECTOR, conditional independence assumptions were made for the purely *ad hoc* and pragmatic reason of allowing the updating to become possible. For Bayesian networks, however, the causal information obtained from the expert provides a justification for making a set of conditional independence assumptions in the manner first suggested by Reichenbach. Moreover, as Pearl, Lauritzen and Spiegelhalter have shown, this set of conditional independence assumptions is sufficient to allow Bayesian updating to become computationally feasible. Everything fits together in a most satisfying manner. There is only one weak link in the chain. It turns out, as we shall see in the next section, that it is

possible to have a *bona fide* causal network in which the corresponding conditional independence assumptions are not satisfied.

4. IMPLICATIONS FOR THE BAYESIAN CONTROVERSY

The preceding sections have outlined some remarkable developments in AI. Let us now turn to a consideration of the implications of these developments for the Bayesian controversy. The Bayesianism versus non-Bayesianism debate has continued among philosophers of science for the last fifty years with no signs of abating. In the 1950s, the major contenders were Carnap (in favour of Bayesianism) versus Popper (against Bayesianism). In the late 1980s and 1990s, we have had Howson & Urbach (1989)[22] in favour of Bayesianism, and Miller (1994)[23] against, while the most recent developments in the debate as seen by leading experts in the area are to be found in Corfield and Williamson, 2001.[24] The new results in AI are clearly relevant to this controversy, and indeed would seem to favour the Bayesian camp, though, as we shall see, this support is more qualified than might at first appear.

The Bayesian controversy, so I believe, involves two rather different issues. The first of these issues is the question of whether we should use the standard mathematical calculus of probability in handling uncertainty, or whether some other calculus might be appropriate. Here, of course, the Bayesians favour the use of the standard calculus. As an example of a non-Bayesian position, we can take the view of Popper [see his (1934)[25], and, for a discussion, Gillies (1998)[26]] that the corroboration of universal laws of science [C(H, E)] is not a probability function, i.e., does not satisfy the standard axioms of probability. In symbols, the claim is that $C(H, E) \neq P(H \mid E)$.

As we have seen, this debate occurred also in the AI context. MYCIN used a non-probabilistic measure of evidential strength, and several other non-probabilistic approaches were proposed and developed by AI workers. [For some details, see Ng & Abramson (1990).[27]] However, the development of AI has given a relatively unequivocal verdict. Probabilistic measures have proved much more successful in practice than non-probabilistic measures, and the latter have tended to disappear. AI has thus supported Bayesianism in this first sense. It should be added, however, that this does not give a decisive verdict against Popper's ideas on corroboration. Popper was considering the corroboration of hypotheses which were universal scientific laws. Most AI systems, however, have as hypotheses singular statements, such as 'this patient's infection is caused by streptococci' or 'that mountain range contains massive sulfide deposits'. It is possible that Bayesianism is appropriate for singular statements, while a non-Bayesian approach is appropriate for universal hypotheses, as I have argued in an earlier work.[26]

Let us now turn to the second and rather different issue involved in the Baysian controversy. It can be most easily approached by considering the form

that the debate has taken within statistics. Classical statisticians such as Neyman were strongly opposed to Bayesianism. Yet Neyman never used any formal system other than the standard mathematical theory of probability. Neyman was clearly not an anti-Bayesian in the sense we have just considered. In what sense, then, was he against Bayesianism? The answer is not immediately clear, since, because he accepted standard probability theory, Neyman *a fortiori* accepted Bayes theorem. The answer to this conundrum is that the second issue in the Bayesian controversy is really about the interpretation of probability. Neyman, following von Mises, regarded the objective interpretation of probability as the only valid one.[28] This meant that some applications of Bayes theorem were illegitimate in his eyes because they necessitated giving a degree of belief interpretation to some of the probabilities used. This applied particularly to the case of giving an *a priori* distribution to a fixed, but unknown, parameter θ. Since θ is fixed and does not vary randomly, it does not make sense to assign it an objective probability distribution, but, since it makes perfect sense for someone to have different degrees of belief in different possible values of θ, θ can easily be given a subjective probability distribution. Many Bayesian analyses involve giving *a priori* distributions to parameters such as θ, and so become illegitimate to a strict objectivist such as Neyman. To sum up: the second issue involved in the Bayesian controversy is really about the relative merits of subjective versus objective interpretations of probability.

What have the AI developments given above shown as regards this controversy? It is immediately clear that they have lent support to the subjective interpretation of probability. Pearl has always argued for a subjective degree of belief interpretation of the probabilities in Bayesian networks, and this remains true of his latest, highly interesting, paper on the foundations of the subject (Pearl, 2001).[29] In this paper, he describes himself as 'only half-Bayesian'. However his departure from standard Bayesianism arises because he thinks that prior probability distributions are inadequate to express background knowledge, and that one needs also to use causal judgements which cannot be expressed in probabilistic terms. As far as the interpretation of probability is concerned, he remains faithful to the subjective degree of belief view, which he says he adopted in 1971 after reading Savage.

Lauritzen and Spiegelhalter were also working in the tradition of subjective Bayesianism, but they seem less definitely committed to this view than Pearl. This is what they say:[30]

> Our interpretation of probabilities is that of a subjectivist Bayesian .... This seems a convenient and appropriate view in an area concerned with the rational structuring and manipulation of opinion, and the subjectivist objectives of a coherent system of probabilities representing belief in verifiable propositions, successively updated on the basis of available evidence, appears to fit remarkably the objectives of expert systems research. However, many of the techniques presented here are appropriate in disciplines where graphical structures are used and a frequentist interpretation is more appropriate, such as complex pedigree analysis in genetics.

So Lauritzen and Spiegelhalter think that, in some cases at least, the probabilities in Bayesian networks might be given an objective interpretation. Neapolitan (1990)[31] is also favourable to objective probabilities in Bayesian networks. So although Bayesian networks were created within the tradition of subjective Bayesianism, it might nonetheless be possible to interpret the probabilities they contain objectively. Arguably, this is likely to be a good strategy in many cases.

A first argument in favour of an objective interpretation is an appeal to the results, mentioned earlier, of de Dombal's group at Leeds. De Dombal and his group devised a computer-based diagnostic system using a straightforward statistical approach. The probability of a patient's having a particular condition given a set of symptoms was calculated using Bayes theorem, where the probabilities employed had been estimated from a large sample of previous patients. This approach did not attempt to encode any medical knowledge beyond what was contained in observed frequencies, and so was not an expert system. As we shall see, however, the system worked very well, and this was because it was designed for use in the following relatively simple situation.

Patients were admitted to the department of surgery of a Leeds hospital because of the onset of acute abdominal pain. The problem was to diagnose the cause of their pain, and, in particular, decide whether an operation was necessary. Leaving aside a small 'other' or 'miscellaneous' category comprising less than 4% of cases, the patients were diagnosed as having just one of the following seven conditions: appendicitis (just over a quarter of the cases), cholecystitis, small bowel obstruction, perforated duodenal ulcer, pancreatitis, diverticular disease, and, last but not least, non-specific abdominal pain, which accounted for about half the cases. The last condition is of course not a disease, but covers those cases where no cause for the pain could be found. This situation is obviously suitable for statistical treatment because of the limited number of mutually exclusive possibilities, and also because during the course of the treatment, often an operation, the cause of the pain could in most cases be definitely established, thereby establishing the correctness or incorrectness of the initial diagnosis.

The efficiency of de Dombal *et al.*'s computer system was compared with that of the hospital clinicians in a sample of 304 patients admitted between 1 January 1971 and 1 December 1971. The overall result[32] was that the computer system was correct in 91.8% of the cases, and the senior member of the clinical team handling the patient in 79.6% of the cases. Some of the detailed differences are also interesting. One of the most difficult, but at the same time crucial, problems in this area is to distinguish between appendicitis and non-specific abdominal pain. If a case of appendicitis is wrongly classified as non-specific abdominal pain, the result could be a delay in operating which results in the appendix perforating or forming an abscess. Conversely, however, if non-specific abdominal pain is wrongly diagnosed as appendicitis, this could result in the patient going through an entirely unnecessary operation – known in the business as a 'negative laparotomy'.

De Dombal *et al.* describe the relative performance in this area of the computer and the humans as follows:[33]

> ... the computer system accurately classified 84 out of a possible 85 patients with acute appendicitis, ... This contrasts with the clinicians' performance, where only 75 diagnoses of appendicitis were made, and six patients were originally classified as non-specific abdominal pain. ... Moreover, although the computer erroneously classified six non-specific abdominal pain patients with the 'appendicitis' category, the corresponding figure for the clinical team was no fewer than 27 patients. ... Had we slavishly followed the computer's predictions, six negative laparotomies would have been performed, but in no case of appendicitis would surgery have been delayed. What actually happened was rather different. Twenty-odd negative laparotomies were performed, and six cases of appendicitis were 'observed' for over eight hours before the decision to operate was taken.

These results are very interesting, but the next experiment of the de Dombal group was perhaps even more interesting (cf. Leaper et al, 1972)[5]. Once again, the trial involved patients admitted to the professional surgical unit of the Leeds General Infirmary with the onset of acute abdominal pain. The period covered again began on 1 January 1971, but this time continued longer until 31 May 1972 producing a larger number (472) of patients. Once again a comparison was made between the diagnoses of the computer system whose probabilities had been calculated from data obtained from a large number of previous patients, and the diagnoses of the senior clinician in charge of the patient. This time, however, a new comparison was introduced. As well as obtaining the probabilities from data, a variant of the computer system was produced in which the probabilities were obtained from estimates provided by the clinicians. In fact, estimates were obtained from six clinicians, and the average taken. The general results of this trial were as follows:[34]

> In the total series of 472 cases the overall accuracy of diagnoses made 'on the spot' by the clinical team was 79.7%, whereas the accuracy of the computer-aided system using values based on 600 surveyed cases was considerably higher (91.1%). ... the overall accuracy of the computer-aided system using the clinicians' estimates was a relatively unimpressive 82.2%.

Once again the results in specific disease categories are also interesting:[35]

> ... in some instances (appendicitis, non-specific abdominal pain) the computer using estimates was more effective than the unaided clinician, but in others (diverticulitis, pancreatitis) it was much less effective. ... the effectiveness of the computer using estimates *seemed to be related to the incidence of the diseases under study*. In respect of acute appendicitis (121 cases in 15 months) and non-specific abdominal pain (230 cases) the computer using estimates was relatively effective when compared with unaided clinician. But for other diseases such as diverticulitis (10 cases) and pancreatitis (14 cases) the computer using estimates proved to be less reliable.

In effect, the computer system using objective probabilities obtained from data outperformed the human clinicians by a considerable margin, but if subjective

probabilities obtained from these clinicians were used instead of the objective probabilities, this gain was wiped out, and in the case of relatively rare diseases the computer system performed worse than the human clinicians. Human doctors appear to be bad at estimating probabilities of diseases, and especially bad in the case of diseases which occur infrequently. It seems hard to avoid the conclusion of the de Dombal group which was the following:[35]

> We suspect that the use of clinicians' estimates of probability may have been a cause of failure in some previous computer-aided diagnostic systems, and we conclude that in future computer-aided diagnostic systems there is no alternative to using carefully collated data from large-scale, real-life surveys rather than clinicians' estimates.

These results are very striking, but the issue is not simply one of a choice between different ways of interpreting probabilities. It should now be pointed out that this choice carries with it methodological implications. If we are interpreting the probabilities as objective, then any proposed value of a probability must be seen as a conjecture which could be right or wrong, and may therefore be in need of testing. Thus, objective probabilities lead to a Popperian methodology of conjectures and refutations in which testing plays a central role. This is indeed the methodology of classical statistics.

Let us next contrast this with the use of subjective probabilities. Any such probability expresses the degree of belief of an individual at a particular moment. Further evidence does not refute the claim that that individual held that degree of belief at that time. It may, however, lead the individual to change his or her degree of belief in the light of the new evidence. In the Bayesian approach, the belief change takes place through Bayesian conditionalisation or updating, i.e., through the change from a prior probability P(H) to a posterior probability P(H | E). In summary, the use of objective probabilities goes with the Popperian methodology of statistical testing; while the use of subjective probabilities goes with the methodology of Bayesian conditionalisation. There have been examples which show that the use of a testing methodology can be advantageous in the construction of Bayesian networks.

One such example [see Sucar (1991)[36], and Sucar, Gillies, and Gillies (1993)[37]] concerned a medical instrument called an 'endoscope'. This allowed a doctor to put into the colon of a patient a small camera which transmitted an image of the interior of the colon to a television screen. In this image an expert could recognise various things in the interior of the colon. Let us take two such things as examples. One is called the 'lumen' which is the opening of the colon. Despite its name, it generally appeared as a large dark region; but sometimes it was smaller and surrounded by concentric rings. Another is called a 'diverticulum' and is a small malformation in the wall of the colon, which can cause some illnesses. A diverticulum generally appeared as a dark region, smaller than the lumen, and often circular. It was a problem then to program a computer to recognise from the image the lumen or a diverticulum. This is a typical problem

of computer vision. To solve it, an attempt was made to construct a Bayesian network with the help of an expert in medical endoscopy.

Figure 5 shows only a small part of this network, but it is sufficient to illustrate the points which are to be made. L stands for the lumen which causes a large, dark, region (LDR) to appear on the screen. This in turn produces values for the variables S (size of the region), and M and V (mean and variance of the light intensity of the region).

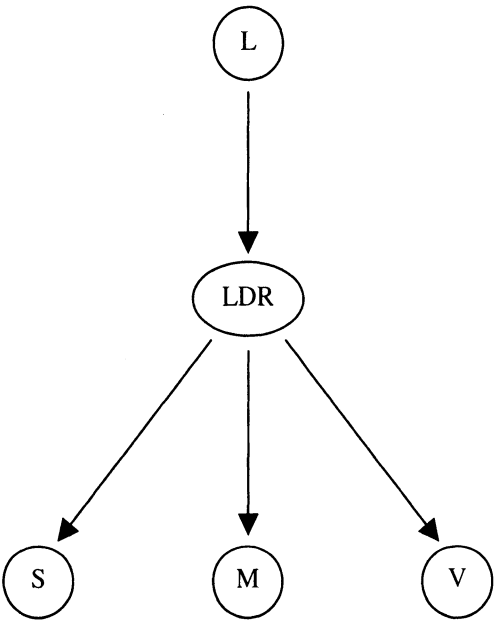


Figure 5

In every Bayesian network, certain assumptions of independence or conditional independence are made. In this case, S, M, V must be conditionally independent of L and mutually conditionally independent, given LDR. Using a Popperian testing methodology, these assumptions were considered as conjectures which needed to be checked by statistical tests. These tests showed, however, that the conditional independence assumptions were not satisfied. In fact, it turned out that, given LDR, M and V were strongly correlated rather than independent.

The response to this situation was to eliminate one of the two parameters M and V on the grounds that, since they were correlated, only one could give almost as much information as both. The results of this elimination were tested using a random sample of more than 130 images of the colon. It turned out that the elimination of one of the parameters gave better results than those obtained using all three parameters. For example, using all three parameters (S, M, V), the

system recognised the lumen correctly in 89% of cases, while, if it was modified by eliminating M, and using only (S, V), it recognised the lumen correctly in 97% of cases.[38] At first sight this seems a paradox, because these better results were obtained using less information. The explanation is simple, however. Undoubtedly there is more information in all three parameters (S, M, V) than in only two (S, V). But the greater amount of information in the three parameters was used with mathematical assumptions of conditional independence which were not correct. The lesser amount of information in the two parameters S, V was, by contrast, used with true mathematical assumptions. So less information in a correct model worked better than more information in a mistaken model. Moreover, since the modified Bayesian network was simpler, the calculations using it were carried out more quickly. So, to conclude, the modified Bayesian network was more efficient, and gave better results. This shows the value of using objective probabilities, and a Popperian methodology of statistical testing.

The example also shows that it is possible to obtain a causal network from an expert for which the assumptions of conditional independence are not satisfied. Thus, although causal networks are useful heuristic guides for the construction of Bayesian networks, they are not infallible guides.

The conclusion I want to draw is that we might move towards a kind of synthesis between the Bayesian and non-Bayesian positions. For subjective Bayesians such as De Finetti, Lindley and Savage, the use of Bayes theorem to update beliefs in the light of evidence was a natural, indeed central, procedure. In the context of expert systems research, the adoption of this procedure led through PROSPECTOR and Pearl's work to the concept of Bayesian network. It is very unlikely that any of the classical statisticians who emphasized objective probabilities and statistical testing would have taken this path. Although such classical statisticians would have accepted Bayes theorem as a consequence of the Kolmogorov axioms, it would not have been natural for them to think of using it, even with objective probabilities, for the purpose of updating in the light of evidence.

Nonetheless, once Bayesian networks had emerged from the programme of subjective Bayesianism, it became clear that such networks could be improved in many cases by incorporating some ideas from classical statistics. These were (1) the use of objective rather than subjective probabilities, and (2) the use of statistical tests to check the assumptions underlying Bayesian networks. In particular, though causal relations in a network do suggest that the corresponding conditional independence relations are satisfied, this may not hold in some cases. Thus, it could be worth using statistical tests to check whether the conditional independence assumptions, considered as conjectures, really hold. If it turns out that they do not hold, this could lead to modifications and improvements in the Bayesian network.

NOTES

1. Jackson, P., *Introduction to expert systems*. Wokingham / England: Addison-Wesley 1986.
2. E.H. Shortliffe & B.G. Buchanan, "A model of inexact reasoning in medicine", in: *Mathematical Biosciences*, 23, 1975, pp. 351-79. The quotation is from p. 357.
3. *Ibid.*, pp. 352-5.
4. F.T. de Dombal, D.J. Leaper, J.R. Staniland, A.P. McCann, & J.C. Horrocks, "Computer-aided diagnosis of acute abdominal pain", in: *British Medical Journal*, 2, 1972, pp. 9-13.
5. D.J. Leaper, J.C. Horrocks, J.R. Staniland, & F.T. de Dombal, "Computer-assisted diagnosis of abdominal pain using 'estimates' provided by clinicians", in: *British Medical Journal*, 4, 1972, pp. 350-4.
6. *Ibid.*, p. 358.
7. J.B. Adams, "A probability model of medical reasoning and the MYCIN model", in: *Mathematical Biosciences*, 32, 1976, pp. 177-86.
8. D. Heckerman, "Probabilistic interpretations for MYCIN's certainty factors", in: L.N. Kanal / J.F. Lemmer (Eds.), *Uncertainty in Artificial Intelligence*, Amsterdam: North-Holland 1986, pp. 167-96.
9. J. Glaschnig, "Prospector: an expert system for mineral exploration", in: D. Michie (Ed.), *Introductory readings in expert systems*, New York: Gordon and Breach 1982, pp. 47-64.
10. R.O. Duda, P.E. Hart, & N.J. Nilsson, "Subjective Bayesian methods for rule-based inference systems", in: *Proceedings of the National Computer Conference (AFIPS)*, 45, 1976, pp. 1075-82. The quotation is from p. 1076.
11. J. Pearl, "Bayesian networks: a model of self-activated memory for evidential reasoning", in: *Proceedings of the Cognitive Science Society*, Ablex, 1985b, pp. 329-34. The quotation is from p. 330.
12. J. Pearl, "Fusion, propagation and structuring in belief networks", in: *Artificial Intelligence*, 29, 1986, pp. 241-88. The quotation is from pp. 253-4.
13. J. Pearl, "Belief networks revisited", in: *Artificial Intelligence*, 59, 1993, pp. 49-56.
14. R.A. Howard & J.E. Matheson, "Influence diagrams", in: R.A. Howard / J.E. Matheson (Eds.), *The principles and applications of decision analysis*, Menlo Park/CA: Strategic Decisions Group 1984, Vol.2, pp. 721-62.
15. J. Pearl, "Reverend Bayes on inference engines: a distributed hierarchical approach", in: *Proceedings of the national conference on AI (ASSSI-82)*, 1982, pp. 133-6.
16. J.H. Kim & J. Pearl, "A computational model for combined causal and diagnostic reasoning in inference systems", in: *Proceedings of the 8th international joint conference on AI (IJCAI-85)*, 1983, p. 190-3.
17. J. Pearl, "How to do with probabilities what people say you can't", in: *Proceedings of the second IEEE conference on AI applications*, Miami/Fl., 1985a, pp. 6-12.
18. J. Pearl, *Probabilistic reasoning in intelligent systems. Networks of plausible inference*. San Mateo/California: Morgan Kaufmann 1988.
19. S.L. Lauritzen & D.J. Spiegelhalter, "Local computations with probabilities on graphical structures and their application to expert systems (with discussion)", in: *Journal of the Royal Statistical Society B*, 50, 1988, pp. 157-224.
20. J. Pearl, "Belief networks revisited", *loc. cit.* (note 13), p. 52.
21. H. Reichenbach, *The direction of time*. Berkeley: University of California press 1956.
22. C. Howson & P. Urbach, *Scientific reasoning; the Bayesian approach*. La Salle/Illinois: Open court 1989.
23. D. Miller, *Critical rationalism*. Chicago and La Salle/Illinois: Open court 1994.
24. D. Corfield / J. Williamson (Eds.), *Foundations of Bayesianism*. Dordrecht/Boston/London: Kluwer 2001.
25. K.R. Popper, *The logic of scientific discovery*. 6th impression of English translation of the original German 1934 edition. London: Hutchinson. 1972.

26. D.A. Gillies, "Confirmation theory", in: D.M. Gabbay / P. Smets (Eds.), *Handbook of defeasible reasoning and uncertainty management systems*, Dordrecht/Boston/London: Kluwer 1998, Volume 1, pp. 135-67. The arguments for Bayesianism for singular statements and non-Bayesianism for universal statements are on pp. 154-5.
27. K. Ng & B. Abramson, "Uncertainty management in expert systems", in: *IEEE Expert*, 5, 1990, pp. 29-48.
28. D.A. Gillies, "Debates on Bayesianism and the theory of Bayesian networks", in: *Theoria*, 64, 1998, pp. 1-22.
29. J. Pearl, "Bayesianism and causality, or, why I am only half-Bayesian", in: D. Corfield / J. Williamson (Eds.), *Foundations of Bayesianism*. Dordrecht/Boston/London: Kluwer 2001, pp. 19-36.
30. Lauritzen & Spiegelhalter, "Local computations with probabilities on graphical structures and their application to expert systems", *loc. cit.* (note 19), p. 159.
31. R.E. Neapolitan, *Probabilistic reasoning in expert systems. Theory and algorithms*. New York: John Wiley 1990.
32. de Dombal *et al.*, "Computer-aided diagnosis of acute abdominal pain", *loc. cit.* (note 4), p. 9.
33. de Dombal *et al.*, "Computer-aided diagnosis of acute abdominal pain", *loc. cit.* (note 4), p. 2.
34. Leaper *et al.*, "Computer-assisted diagnosis of abdominal pain using 'estimates' provided by clinicians", *loc. cit.* (note 5), pp. 351-2.
35. Leaper *et al.*, "Computer-assisted diagnosis of abdominal pain using 'estimates' provided by clinicians", *loc. cit.* (note 5), p. 353.
36. L.E. Sucar, *Probabilistic reasoning in knowledge-based vision systems*. PhD thesis, Imperial College, University of London 1991.
37. L.E. Sucar, D.F. Gillies, & D.A. Gillies, "Objective probabilities in expert systems", in: *Artificial Intelligence*, 61, 1993, pp. 187-208.
38. *Ibid.*, p. 206 for further details.

Department of Philosophy
King's College London
Strand
London WC2R 2LS
U.K.
donald.gillies@kcl.ac.uk

STEPHAN HARTMANN

ARTIFICIAL INTELLIGENCE AND ITS METHODOLOGICAL IMPLICATIONS

Comments on Donald Gillies's "Handling Uncertainty in Artificial Intelligence, and the Bayesian Controversy"

1. INTRODUCTION

Donald Gillies is one of the pioneers in the philosophical analysis of artificial intelligence (AI). In his recent book, Gillies (1996) not only makes a new and rapidly developing field of science accessible to philosophers; he also introduces philosophical topics relevant to researchers in AI and thereby helps establish a dialogue between the two disciplines. His book clearly and convincingly demonstrates the fruitful interplay between AI and philosophy of science.

The present paper continues the discussion in the book by focusing on methodological issues. I agree with a lot of what Gillies has to say, but commentaries tend to be more interesting if the commentator focuses on controversial issues. So I will discuss three topics on which our views differ. First, I reconsider the issue of the relevance of scientific developments for a philosophical debate (Sec. 2). More specifically, I will address the following question: Do we have to change our philosophical views in the light of the development of new scientific *tools* (such as Bayesian Networks)? Contrary to Gillies, I will argue that the new developments in AI have no impact on what Gillies calls the Bayesian controversy. Second, I discuss two of Gillies' concrete methodological recommendations. Gillies suggests that the use of objective probabilities and the construction of a Bayesian Network model should be supplemented by a Popperian testing methodology. I will argue that the process of finding the value of an objective probability is not an instance of a test (Sec. 3), and that there are alternative (and perhaps more efficient) strategies for the construction of a Bayesian Network model (Sec. 4). Third, I am concerned with the scope of Bayesianism. Gillies gives an argument to the effect that Bayesianism does not apply to general hypotheses. I will show that this argument does not go through (Sec. 5).

F. Stadler (ed.), Induction and Deduction in the Sciences, 217–223.
© 2004 *Kluwer Academic Publishers.*

2. THE PHILOSOPHICAL IMPLICATIONS OF AI

There is no doubt that science has an effect on our philosophical views. Some scientific developments have inspired work in metaphysics (e.g. quantum statistics gave rise to a reconsideration of the concept of individuality), and other developments forced philosophers to rethink their positions (e.g. about absolute space). The situation in AI, however, is different. Here, a new scientific tool is developed (i.e. Bayesian Networks) and Gillies claims that this development has methodological implications for the Bayesian controversy. This controversy is concerned with two issues. The first is about the use of probabilities (as opposed to other ways of representing uncertainty in AI). The second is the interpretation of probability. Let me examine both issues in turn.

A survey of the literature reveals that the majority of approaches to AI use probabilities, while only a minority deals with alternatives such as the Dempster-Shafer theory, ranking functions, or Popper functions. This raises the question of whether the calculus of probabilities is appropriate to handling uncertainty. Probabilities are very popular in scientific practice. But can we conclude from this that probabilities are *better* than alternative options; i.e. can we conclude from this that the right way to represent uncertainty is by using probabilities? I do not think so. The popularity of Bayesian Networks in the AI community can be understood by noticing two points. First, there is an extremely fruitful research program centered around Bayesian Networks which gives researchers in the field a chance to make their own contributions to the development of the theory. Second, Bayesian Networks are very easy to apply. In fact, they are much easier to apply than any of the other available tools, which is partially rooted in the fact that the alternatives to Bayesian Networks are simply not developed well enough yet. These pragmatic factors play an important role when a decision has to be made as to which tool we choose to work with. However, these pragmatic factors do not support any deeper philosophical claim.

When discussing non-Bayesian approaches of handling uncertainty, Gillies himself does not refer to alternatives such as the Dempster-Shafer theory, but to measures of confirmation C(H, E) for a hypothesis H by evidence E that do not satisfy the standard axioms of probability. In symbols, the claim is that $C(H, E) \neq P(H|E)$. This is confusing because different measures of confirmation (Fitelson 1999) have been discussed extensively within the Bayesian literature. These measures (such as the difference measure, the ratio measure, and the log-likelihood measure) do not satisfy the axioms of probability theory. Gillies also mentions the certainty factors used in the expert system MYCIN, which do not obey the axioms of probability theory. It should be noted, however, that Heckerman (1986) shows that a consistent probabilistic interpretation of certainty factors *can* be given. According to this analysis, certainty factors are monotonic transforma-

tions of the likelihood ratio $P(H|E)/P(H|\neg E)$. This demonstrates that there are many different ways of handling uncertainty also within the Bayesian paradigm.

The second aspect of the Bayesian controversy is the interpretation of probability. Gillies suggests that the development of Bayesian networks supports a subjective interpretation because they were developed within the tradition of subjective Bayesianism. I do not think that this argument holds. To see why, consider the following analogy. Quantum mechanics in its current form was developed within the tradition of the Copenhagen approach of quantum mechanics. Bohr, for example, entertained the ideas that later became part of the Copenhagen interpretation long before the development of the formal theory in the 1920s. And yet, no one would claim that the success of quantum mechanics provides an argument for this interpretation. Later, other interpretations (like the many-worlds or the many-minds interpretation) were developed which turned out to be completely compatible with quantum mechanics. And indeed, Gillies himself points out that Bayesian Networks can also be applied if the probabilities in question are objective. Hence, even if Bayesian Networks were created within the tradition of subjective Bayesianism, the theory of Bayesian Networks itself is *neutral* with respect to the interpretation of probability (Bovens and Hartmann 2002, Williamson 2001). In the first place, Bayesian Networks are a highly efficient tool to represent and manipulate probabilistic structures, irrespective of how the probabilities are interpreted. One might reply that Bayesian Networks require the specification of some prior probabilities which is, after all, a typical Bayesian step. This, however, is not correct since Gillies' own examples from medicine show that prior probabilities can also be objective.

Let us now examine if the subjective interpretation of probability was heuristically important for the development of Bayesian Networks. In his personal account, Pearl (1993) mentions a lot of factors that played a role there, but he does not mention subjective probabilities. Apart from paying lip service to subjective Bayesianism, no deeper influence of this methodological account is acknowledged. Apparently the Bayesian framework is flexible enough to allow for an incorporation of the technical ideas and practical applications that guided Pearl's thinking. In any case, philosophical questions concerning the interpretation of the probability calculus do not seem to have been that important after all.

To sum up, I do not see that the development of Bayesian Networks has an effect on the Bayesian controversy in philosophy of science. The question whether Bayesianism is a good methodology or not is of course important. It is, however, not decided or even illuminated by pointing to a commonly used tool in AI.

3. OBJECTIVE PROBABILITIES

Gillies makes several methodological recommendations in his paper. He argues, for instance, that objective probabilities should be used whenever possible and

that the use of objective probabilities requires a Popperian methodology. Let us examine these two claims. To support the first claim, Gillies discusses an example from medical diagnostics. It turns out that expert systems based on objective probabilities yield better results than systems that use averaged subjective probabilities provided by experts. I agree, but I do not find this particularly astonishing. After all, the subjective probabilities provided by the experts are nothing but guesses of the values of objective probabilities. These guesses are based on the experience of the expert, on the experience of their colleagues, and, perhaps, on what the experts took from textbooks and journal articles. That is, the subjective probabilities discussed here are hypotheses about the values of objective probabilities. An expert estimates these probabilities and this, of course, may involve errors and inaccuracies. The error will be expected to be particularly high if the sample an expert uses for his judgements is small. And this is just what Gillies' examples show.

Let us now turn to Gillies' second claim. Do objective probabilities require a Popperian testing methodology? Gillies argues that the assignment of an objective probability is a conjecture, and conjectures have to be tested. If the hypothesis does not pass the test, it is falsified and a new conjecture has to be made. I doubt that this reconstruction of scientific practice is adequate. True, to claim that an objective probability has a certain value is a scientific hypothesis. This claim is based on the available data. If new data become available, the corresponding value is *updated* (and not revised), i.e. it is changed according to all available evidence. This is done on the basis of the new *and* old data jointly. Such an updating procedure does not correspond to a test of the original hypothesis, which was, perhaps, falsified. The original hypothesis was simply modified on the basis of more data.

It is worth pointing out that such a procedure is not a problem for a Bayesian. The modification of the value we claim an objective probability to have is a different procedure than the updating of the probability of a hypothesis in the light of new evidence. It can be done "by hand" without using Bayes' theorem (Hawthorne 1993, p. 134).

4. A TESTING METHODOLOGY FOR BAYESIAN NETWORKS?

According to Gillies, a testing methodology is not only needed for the specification of various objective probabilities, but also for the determination of the structure of a Bayesian Network. In his example from medicine, one starts with causal knowledge from experts. On the basis of this knowledge, a Bayesian Network model is constructed. Now, this network implies various probabilistic independencies which may turn out not to be reflected in the statistical data the network is supposed to represent. In this case, the hypothesis that the network in question represents the statistical data is falsified and a new hypothesis has to be found. Hence, Gillies suggests the following methodology to construct a Bayes-

ian Network model: (1) start with the causal knowledge of experts and construct a Bayesian Network; (2) use all available data to test the independence assumptions (and perhaps also other assumptions) made by the network model. These tests are attractive because they provide an answer to a yes/no question: an independence condition either holds or it does not. (It is assumed that the values of the objective probabilities were already tested.) (3) If the model does not pass a test, a modified network has to be suggested based on new knowledge about various independencies.

In which situations is this procedure a good methodology? Certainly if there is reliable causal knowledge and if the number of variables is small. Both conditions hold in Gillies' example from medicine. Moreover, they are often related, for it is harder to formulate reliable causal claims if the number of variables is large. However, if the number of variables is large, there may not be enough computational power available to test all independencies represented by the network.

What alternatives are there to Gillies' Popperian methodology? I will discuss three. First, if there is no causal knowledge to start with, a methodology based on Bayesian Networks can be used to discover causal knowledge. Algorithms, such as the ones developed by the CMU group, aim at finding a Bayesian Network which fits a certain set of statistical data best (Spirtes et al. 2000). An additional Popperian testing methodology is not needed here. It should be mentioned, however, that this methodology has a lot of problems of its own and so far not much causal knowledge has been gained.

A second alternative is to apply a Bayesian methodology. Here, tests are also possible. If the data is inconsistent with the independence assumptions represented by the network, the probability of the data given the network, P(data| graph), equals zero and hence, by Bayes' Theorem, we have P(graph|data) = 0. This in turn leads to a falsification of the proposed network model (Howson and Urbach 1996, p. 84f). A problem with this procedure is that it does not directly identify the faulty independence assumption. However, given additional background knowledge, a negative test may suggest a better model.

A third strategy, which is also a Bayesian strategy, is to start with several possible candidate models and then eliminate them successively when new data become available. Hawthorne (1993) defended this methodology and proved a number of limiting theorems that aim at demonstrating the practicability of this procedure. This strategy does, however, not seem to be realistic given certain computational limitations.

To sum up, there seems to be a plurality of strategies which guide the construction of a Bayesian Network. Gillies' Popperian testing methodology is one of them. All of these strategies have problems. It will be interesting to learn more about the conditions under which the various methodological strategies can be successfully applied.

5. THE SCOPE OF BAYESIANISM

Gillies mentions in his paper, and argues in more detail elsewhere, that he accepts Bayesianism only if it is about specific hypotheses. He presents a new argument for the conclusion that general hypotheses have probability zero (Gillies 1998, p. 154f). This implies, by Bayes Theorem, that also the posterior probabilities of general hypotheses are zero which, in turn, renders the whole Bayesian program useless in these cases. I will argue that Gillies' argument is flawed.

The argument is based on the observation that one cannot win a bet on a general hypothesis. One can only lose it. Hence one should not bet on a general hypothesis, which implies that its probability is zero. Let us look at a specific case. Let H be the hypothesis "All ravens are black". Since H is a general hypothesis, $P(H) = 0$ and so $P(\neg H) = 1$. Hence we should be prepared to bet *everything* on the hypothesis that there is a non-black raven. I am not prepared to do so, and I am not sure if Gillies is. I think that my argument shows that something is wrong with Gillies' de Finetti style argument. I take it to be much more plausible to follow the programme outlined by Ramsey and to assign probabilities on the basis of utilities. This might indeed lead to a non-zero probability for a general hypothesis.

Having defended Bayesianism against the charges of Gillies, I should mention a problem which Bayesians and Popperians share. Both methodologies assume that a decision can always be made. For a Popperian, a hypothesis either passes a test or is falsified; for a Bayesian it can always be decided which hypothesis of a set of hypotheses should be accepted. This might not always be possible. One can imagine situations in which there is not enough evidence to make a decision because there is too much uncertainty. Instead of making a definite decision, scientists should withhold judgement and try to collect more relevant information. That is, I suggest to weaken the link between Bayesianism and decision theory.[1]

To sum up, I have argued in this paper that the remarkable developments in AI have no baring on the Bayesian controversy. New scientific tools such as Bayesian Networks do not provide arguments in favor of a specific interpretation of probability or a specific scientific methodology. It is up to philosophical analysis to provide arguments in favor or against one of the various options.

Acknowledgements: I would like to thank Donald Gillies, Jim Hawthorne, Jürg Kohlas and Jon Williamson for discussions and Luc Bovens, Roman Frigg and Gabriella Pigozzi for carefully reading this paper.

NOTE

1. I owe this point to Jürg Kohlas.

REFERENCES

Luc Bovens and Stephan Hartmann, "Bayesian Networks and the Problem of Unreliable Instruments", in: *Philosophy of Science* 69, 2002, pp. 29-72.

Branden Fitelson, "The Plurality of Bayesian Measures of Confirmation and the Problem of Measure Sensitivity", in: *Philosophy of Science* 63, 1999, pp. 652-660.

Donald Gillies, *Artificial Intelligence and Scientific Method*. Oxford: Oxford University Press 1996.

Donald Gillies, "Confirmation Theory", in: Philippe Smets (ed.), *Handbook of Defeasible Reasoning and Uncertainty Management Systems. Volume 1, Quantified Representation of Uncertainty and Imprecision*. Dordrecht: Kluwer Academic Publishers 1998, pp. 135-167.

James Hawthorne, "Bayesian Induction Is Eliminative Induction", in: *Philosophical Topics* 21, 1, 1993, pp. 99-138.

David Heckerman, "Probabilistic Interpretations for MYCIN's Certainty Factors", in: Laveen N. Kanal and John F. Lemmer (eds.), *Uncertainty in Artificial Intelligence*, Amsterdam: North-Holland 1986, pp. 167-196.

Colin Howson and Peter Urbach, *Scientific Reasoning: The Bayesian Approach*. Chicago: Open Court 1996.

Judea Pearl (1993), "Belief Networks Revisited", in: *Artificial Intelligence* 29, 1986, pp. 241-288.

Peter Spirtes, Clark Glymour and Richard Scheines, *Causation, Prediction, and Search*. Cambridge, MA: MIT Press 2000.

Jon Williamson, "Foundations for Bayesian Networks", in: David Corfield and Jon Williamson (eds.), *Foundations of Bayesianism*. Dordrecht: Kluwer Academic Publishers 2001, pp. 75-115.

Department of Philosophy, Logic and Scientific Method
London School of Economics and Political Science
Houghton Street
London WC2A 2AE
U.K.

and

Zentrum für den wissenschaftlichen Nachwuchs
Universität Konstanz
78457 Konstanz
Germany
Stephan.Hartmann@uni-konstanz.de

HANS ROTT

SUPPLYING PLANKS FOR NEURATH'S BOAT: CAN ECONOMISTS MEET THE DEMANDS OF THE DYNAMICS OF SCIENTIFIC THEORIES?

1. INTRODUCTION

According to Otto Neurath, the practice of science consists in a large undertaking of setting up and maintaining systems of statements:

> In unified science we try ... to create a *consistent system* of protocol statements and non-protocol statements (including laws). When a new statement is presented to us we compare it with the system at our disposal and check whether the new statement is in contradiction with the system or not. If the new statement is in contradiction with the system, we can discard this statement as unusable ('false'), for example, the statement: 'In Africa lions sing only in major chords'; however, one can also 'accept' the statement and change the system accordingly so that it remains consistent if this statement is added. The statement may then be called 'true'.
>
> The fate of being discarded may befall even a protocol statement. There is no '*noli me tangere*' for any statement though Carnap claims it for protocol statements (Neurath 1932/33, pp. 94-95)

Neurath's coherentist epistemology – famously expounded in the *Protokollsatz-debatte* with Moritz Schlick – rejects any commitment to a distinguished set of sentences that might be used as firm foundations upon which to build all of science. The scientist's task is to keep his system coherent, not to pay undue respect to a predetermined class of privileged truths. Neurath was very clear about the fact that there are many consistent systems that might be offered to account for the observed phenomena. In contrast to Schlick, Neurath indeed welcomed the idea that there is no unique "criterion of truth". Scientists have to make decisions which of these systems to adopt, and indeed there is no escape from making decisions in science.

Neurath, however, said disappointingly little about *how* to choose the most adequate system of propositions among multiple coherent candidate systems. In "Sociology in the Framework of Physicalism" (1931, p. 66), he remarked that on encountering some recalcitrant data it is hard to decide in favour of the data and a revision of the system one has been using so far. A little later, in "Radical Physicalism and the 'Real World'" (1934, p. 109), he held that scientists should

F. Stadler (ed.), Induction and Deduction in the Sciences, 225–245.
© 2004 *Kluwer Academic Publishers.*

aim at making their theories agree with as many protocol statements as possible.[1] While Neurath took the latter requirement to be a mark of empiricism, we can also interpret it as an element of conservatism. When trying to accommodate new observations, scientists should never sacrifice the observations that had been successfully accounted for before the new observation was made. This fits together with the general reluctance ("*wozu man sich im allgemeinen schwer entschließt*") of scientists to revise a corroborated system that Neurath was speaking about in his earlier paper.

The textual evidence for this interpretation, however, is thin. It was only Schlick, Neurath's opponent in this discussion, who in his paper "The Foundation of Knowledge" made a clear suggestion as to how the problem of choice involved might be solved. In a way congenial with Neurath, he suggested employing a *principle of economy*:

> By what mark, then, are we to distinguish these statements which themselves remain unaltered, while all others must be brought into agreement with them? ... The most obvious recourse would doubtless be to find the rule for which we are searching in some kind of *economy principle*, namely, to say: we are to choose those as basic statements whose retention requires a *minimum* of alteration in the whole system of statements in order to rid it of all contradictions. ...
>
> It is worth noticing that such an economy principle would not enable us to pick out certain statements as being basic once and for all, ... This would, of course, no longer be the pure *coherence* viewpoint, but one based on *economy* (Schlick 1934, pp. 216-217)

Schlick thought that the proponents of the coherence theory of scientific knowledge he attacked "did in fact take the economy principle as their guiding light, whether explicitly or implicitly", but he argues, against them, that the epistemic merits of a statement consist in its pedigree rather than its pragmatic utility.[2]

The present paper takes a look at what may be viewed as a successor project to Schlick's idea of supplementing Neurath's picture of the structure and dynamics of science by an economy principle. In the 1980s, logical models were developed that explicitly addressed the problem of maintaining consistency in systems of statements, and it was often proclaimed that the particular way of choosing among consistent successor theories was guided by an economy principle very much like Schlick's.[3] I am referring to the *theory of theory change*, which is also known as *belief revision*. The theory has been rather limited, because it postulates that new information be always accepted. In this respect the model falls behind Neurath's (1932/33, pp. 94-95) insight that in case of conflict one might *either* revise the theory *or* reject the new statement as "useless" or "false".[4] However, this seems to me a minor defect, one that does not invalidate the discussion below and that has recently been rectified anyway.[5]

Where does this account leave deduction and induction in the sciences? According to Neurath, inductive reasoning always involves decisions as to which sentences to adopt, while deductive logic serves as a standard against which to measure whether potential theories are free of contradictions or not. *Logic in a*

narrow sense refers to formal models of deductive reasoning, axiomatic systems which should preferably come together with some semantic underpinning. *Logic in a wider sense* is the theory of good reasoning, providing us not only with the means to check candidate theories for consistency but also with a methodology for deciding which of the candidate theories to adopt. Neurath (1934, p. 106) rightly laid stress on the fact that theory choice is based on *extralogical* considerations, but in this he clearly referred to the narrow conception of logic. I will in this paper proceed on the assumption that logic does include rules that are relevant for processes of belief formation and theory choice (a kind of inductive reasoning, see Spohn 2003). We shall see that we can clearly distinguish an idea of *economical* behaviour (or *economy*) from behaviour as recommended by *economics* (*economic* behaviour). We will ask to what extent the two ideas have as a matter of historical fact become embodied in formal models for theory formation and transformation, and we will address the question to what extent they *should* be respected in such models. In doing so, I will hardly say anything about the role logic has to play in (empirical) science, real or reconstructed. This paper is not a piece of formalised philosophy of science. Rather, I want to have a look at the question as to the role that the science of economics may play in logic broadly conceived. In order to do this, we first have to get an idea of what economics and economy are.

2. WHAT IS ECONOMICS?

One way of finding out what a term means is to look at the science that is supposed to study it. For the word 'economic', this seems to be an easy task. We just have to look at what economics is about. Dictionaries define 'economics' as the scientific study of the production, distribution and consumption of goods, services and wealth, or more concretely, the study of the system of trade, industry, money etc. But we want to dig deeper. According to Francis Y. Edgeworth (1881, p. 16), "[t]he first principle of Economics is that every agent is actuated only by self-interest." This restriction to a completely selfish attitude has long been removed. The formal part of Edgeworth's idea, however, remains valid. In the words of Herbert Simon: "The rational man of economics is a maximiser, who will settle for nothing less than the best."[6] This still seems to be the dominant view, as is shown by a few more recent statements. Hausman (1998) gives the following summary of the dominant school in the 20th century: "The main 'orthodox', 'neoclassical', or 'neo-Walrasian' school models economic outcomes as equilibria in which individuals have done as well for themselves as they could given their preferences and the constraints on their choices." In another handbook article, Rosenberg (1995) tells us about the "assumptions of the 'economic man': that all agents have complete and transitive cardinal or ordinal utility rankings or preference orders and that they always choose that available option which maximises their utility or preferences". Earlier the same author character-

ised economics as guided (or misguided) by an "extremal intentional research programme" (Rosenberg 1983).[7] And what is being maximised is utility. According to Broome (1999, p. 21-22): "'utility' acquired the meaning: *the value of a function that represents a person's preferences.* ... The first principle of economics is ... utility theory, ... modern, axiomatic utility theory ..."

What is common to these very abstract formulations of the basic tenets of modern economics is that economic agents are viewed as having definite preferences, and that when choosing actions or commodities, they aim at satisfying their preferences as well as the circumstances allow. It is important to note that not just any kind of preference is considered to be appropriate. Preferences have to be transitive and complete (technically speaking, they must be "preorderings" or "weak orderings"), in order to be representable by a suitably chosen utility function. Ties in the preferences are permitted, of course, but incomparabilities are ruled out. Put very pointedly, one can say that economics is based on (and, in a very abstract sense: is an elaboration of) the theory of rational choice. Here, *choice* is called *rational* or *coherent* if and only if it is representable by a preference relation, and a *preference relation* in turn is considered to be *rational* if and only if it is representable by a *utility function.*[8]

Economics, then, is about *rational* rather than the *actual* behaviour of individuals. The laws of economics should not be expected to be empirically adequate, they are valid only as idealisations, or as norms. Economics is based on a formalised variant of common-place folk psychology with individual-level explanation of free agents. It is essential to the research program of economics proper that processes pertaining to whole societies or economies are to be explained by, or reduced to, the behaviour of egocentric, maximising agents.[9] Usually, it is assumed that comparability of preferences or aggregability of utilities across different individuals does not make sense. It is important, however, that comparisons and aggregation are taken to be possible across different "aspects" or "criteria" of goods, as well as across different potential situations in which the same goods would be available (to various extents). This is part of ordinal and expected utility theory, and I suspect that this may ultimately be the reason for the economists' insistence on a person's preferences being representable by a utility function.

In order to find out whether there is anything economic (referring to economics in this abstract standard sense) about theory revision, we will have to look for patterns of rational choice, i.e., choice that is maximising with respect to some underlying preferences, or more exactly, to some weak ordering which can in turn be represented by a utility function.

3. A SECOND VIEW OF ECONOMICAL BEHAVIOUR: "INFORMATIONAL ECONOMY"

There is an alternative and perhaps more intuitive concept of "acting economically" that has played an important role in the development of the research on theory revision. Theory revision is often thought to be *economical* behaviour rather than *economic* behaviour. In the English language, there is a division of labour between the adjectives "economic" and "economical". The former is closely tied to the noun "economics" and means either "of or referring to economics" or has "profitable", "remunerative" and "gainful" as potential synonyms. The adjective "economical", on the other hand, is closely tied to the noun "economy" and means essentially the same as "thrifty", "frugal" or "not wasteful". This particular difference of meanings turns out to be useful for our discussion.

The perspective of *economy* (rather than *economics*) was forcefully taken in Peter Gärdenfors's enormously influential book *Knowledge in Flux* (1988). A glance at the index of the book makes immediately clear that the *criterion of informational economy* is supposed to motivate the essential part of the formal modellings of Gärdenfors and his collaborators Carlos Alchourrón and David Makinson. Gärdenfors employs this criterion for the motivation of belief expansions (p. 49), belief revisions (pp. 53, 58) and belief contractions (p. 61).[10] Basically, the criterion is taken to be identical with the idea of *minimal change* (p. 53) and the *conservativity principle* (p. 67). This is very close to Schlick's interpretation of Neurath. According to Gärdenfors,

> The key idea is that, when we change our beliefs, we want to retain as much as possible of our old beliefs – information is in general not gratuitous, and unnecessary losses of information are therefore to be avoided. (Gärdenfors 1988, p. 49, similarly on pp. 16, 157)

Ever since the appearance of Gärdenfors's book, the criterion of informational economy has been taken to be a "hallmark" of the paradigm developed by Alchourrón, Gärdenfors and Makinson.[11] There is, however, reason to ask why exactly a rational person should be conservative. Gärdenfors's argument that information is not gratuitous does not seem to be sufficient, because information, even if costly, can be wrong, and even if it is correct, it can be misleading. Spelt out a little more explicitly, the argument for conservatism seems to be this:

> You shouldn't give away what is valuable
> What you have is valuable
> ---
> You shouldn't give away what you have

The first premise is hardly controversial, it might even be called analytically true as a prescription of economic rationality.[12] The second premise, on the other hand, is much harder to substantiate. It (or at least the idea that it is true) may have evolutionary reasons – probably we would not be alive any more if too many of the statements we accept as true were wrong. Therefore it is at least likely that many of the statements that we happen to hold true now have some survival value.[13]

But the argument for conservatism is a far cry from waterproof. Problems are not hard to come by. Isn't it all too obvious that what we have is not always best? So why should we care to preserve it? It is true that we lose *information* or *content* when we give up some statements of our theory, but it is not clear whether we lose some *truths*, and no-one should object to losing falsities. Notice that there is a basic tension here between the economical and the economic precept: The former tends to recommend leaving everything as it is, while the latter recommends striving for the best.

4. ECONOMIC AND ECONOMICAL CONSIDERATIONS IN THE THEORY OF THEORY CHANGE

In asking what is economic about theory change, we have to keep in mind two different aspects. After the choice-preference-utility line of thinking that we sketched in Section 2, we have found a second one in the thrifty adherence to the sentences one has accepted.

In the work of AGM, it is comparatively easy to recognise the criterion of informational economy at work in expansions of theories by statements that do not contradict the prior theory. AGM recommended adding the new statement and taking the deductive closure.[14] However, as we shall see, there are no traces of the criterion for the theory-contravening case which, after all, is the case for which logical models of theory change have primarily been devised. But Gärdenfors's argument can be generalised. In my view, one of the most important philosophical advances of belief revision theory in the last decade is the clear recognition that *belief states* cannot be represented properly by *theories* only. Something else has to be added, namely, some mechanism that tells the agent how to revise her theory. Typically this mechanism exploits some sort of selection function or preference relation.[15] Let us use a neutral name and call the structure exploited by the mechanism a *belief-revision guiding structure*.[16] This argument can be translated into the language of theory change. It then says that scientists do not only subscribe to the theories that they accept, but also to dispositions how to change them should they turn out to be wrong (i.e., roughly, scientists subscribe to a Lakatosian positive heuristics). It turns out that if the new information is inconsistent with the presently accepted theory, it is impossible to apply the pure idea of informational economy on the level of statements. But the idea can be applied on the level of revision-guiding structures. This at the

same time defines a form of conservatism that is usable for iterated theory change. Fig. 1 gives a representation of the many senses that "economic" and "economical" can take in theory revision contexts.

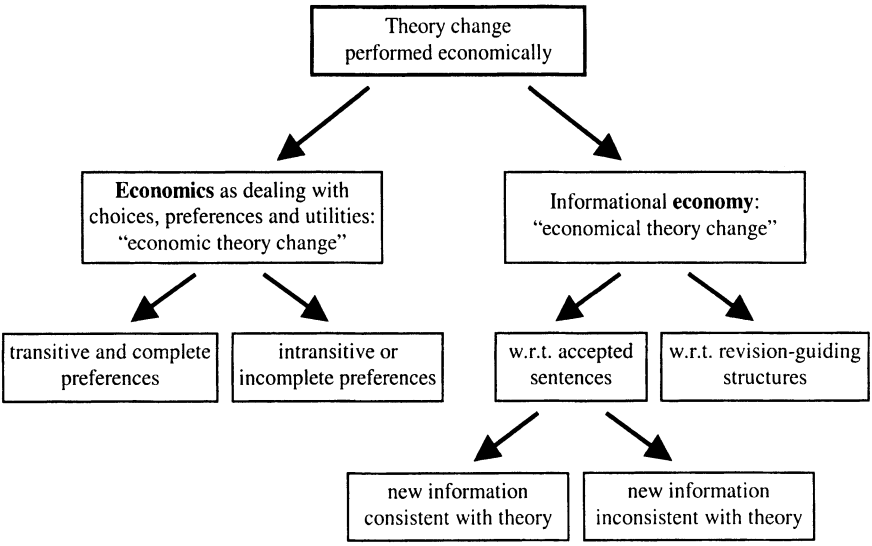


Figure 1

The branch at the left indicates the idea that economic agents are rational or coherent choosers. Though this has not been the principal motivation of theory revision models, we will later see that exactly this idea can be located in the realm of theory formation. On the right-hand branch, we have the idea of informational economy that has always been advertised as the prime driving force of the theory of theory revision (even if thrift in itself does not play a major role in the science of economics). In the next sections, we will trace a few important distinctions within the principle of informational economy, also known as principle of minimal change or conservatism. The maxim of informational economy with respect to sentences tells us: "Don't give away accepted sentences beyond necessity!" The maxim of informational economy with respect to revision-guiding structures tells us: "Don't change your preferences beyond necessity!" These two maxims can further be distinguished as to whether they apply to the theory-contravening case (the one for which theory revision models have originally been invented) or only to the case where the new information is consistent with the agent's theory.

Before entering the systematic discussion, let us summarise some history. A basic form of theory change represents methods that recognise neither economical constraints of minimal change nor economic constraints concerning the rationality of choices. It is possible to add elements of conservatism and ele-

ments of rational choice independently from one another. In their seminal work in the 1980s, AGM provided for pure conservatism with respect to sentences (so-called *maxichoice* contraction and revision functions) and for completely unconstrained choices (the case where only their six *basic* rationality postulates hold).[17] However, they rejected the idea of maxichoice functions as intuitively inadequate, and I think it is fair to say that the main elegance and force of their theories derives precisely from the *supplementary* postulates that go beyond their basic case. So what distinguishes the work of AGM is, on the one hand, a rather strong concept of rational choice generated by transitive and connected preferences, and on the other hand a rather weak concept of conservatism. They provided for conservatism with respect to accepted sentences in the consistent case (where revision reduces to set-theoretic addition plus logical closure), but they provided neither for conservatism in the theory-contravening case nor for conservatism with respect to revision-guiding structures (they provide no change mechanisms for revision-guiding structures at all). So contrary to first appearances, AGM paid a lot more respect to ideas found in economics than to the idea of informational economy.

Systematic variations of the strength of the relevant ideas were investigated only in the 1990s. On the one hand, *weakenings* of AGM's strong presuppositions with respect to the rationalisability of choices are suggested by Lindström (1991) and Rott (1993, 2001).[18] On the other hand, *strengthenings* of the very weak concept of conservatism in AGM are investigated by Boutilier (1993, 1996) and Rott (2003a). Interestingly, it has turned out that the most efficient way – and perhaps the only way – to implement conservatism with respect to accepted sentences in the theory-contravening case is at the same time a form of conservatism with respect to revision-guiding structures. This variant of conservative theory change was actually first discussed as a particular strategy of extending the classical AGM model so as to equip it with means for performing iterated theory changes. While Boutilier did that in a context that presupposes the full strength of AGM's requirements for rational choice, Rott lifts this restriction and shows that the conservative method can be brought to bear without any assumptions about the coherence of the choices involved.

This little sketch of some developments in literature on theory change indicates that the concept of *economical* theory change is independent of the concept of *economic* theory change. In the sense specified, AGM theory revision (minus their two axioms for revisions by inputs that are consistent with the prior theory[19]) is not at all economical but, thanks to their "supplementary" postulates, fully economic; the conservative approach investigated in Rott (2003a) is very economical, but not at all economic.

In the following sections, we will recapitulate relevant parts of the actual development of the theory of theory revision, and then discuss the merits and problems of changing theories economically. I will first argue that conservatism with respect to *accepted sentences* has hardly been followed in classical AGM-style theory revision models, and that it is doubtful indeed whether it would be a

good idea to follow it. Conservatism with respect to theory-revision guiding structures has been studied as *one* strategy among others in iterable theory change, but it has turned out that it should not be followed either. On the economics side, fairly severe constraints on rational choices have been endorsed in (iterable extensions of) classical AGM-style theory revision. It is true that the AGM postulates can be liberalised systematically according to one's wishes. But we will show that even more modest postulates are problematic – just as problematic as even the basic constraints on rational choices are. We shall see that preferences appear to be context-dependent, giving rise in particular to the formidable problem frequently labelled "the informational value of the menu".

5. INFORMATIONAL ECONOMY WITH RESPECT TO THEORIES

In discussing informational economy with respect to theories, we will continue to make two important idealisations. We stipulate that the set of statements accepted by a scientist be logically consistent and closed. This requirement is, of course, wrong as a description of the set of sentences that scientists would affirm to when queried in an interview. However, if we consider a theory to be the set of sentences that we *ascribe* to a scientist from a third person perspective, or the set of sentences that the scientist is *committed to*, then the ideas of consistency and closure lose their implausible appearance. Another simplification we want to make is that when a theory is revised in response to some new piece of information, the revision process in fact successfully incorporates the new information, so that it is accepted in the revised theory. We were taught by Neurath that this is not always reasonable in realistic theory change situations, but I think we can disregard the complications for our discussion. So let us call a consistent and logically closed theory that includes a new piece of information α a *candidate revision* of theory T by α. A (candidate) revision is called *theory-contravening*, if α is inconsistent with the prior theory T.

Let us discuss two attempts at capturing the idea of informational economy on the level of theories:

(1) When accepting a new piece of information, a scientist should aim at a minimal change of his old theory.

(2) If there are different ways of effecting the theory change, the scientist should give up those statements that are least entrenched.

These two maxims have frequently been called upon as the principal motivation for logical models of theory change. However, in their most straightforward readings, they are a caricature of what has really been done in the development of the standard models of theory revision. I have argued for this in Rott (2000), and briefly recall the results presented in that paper.

As regards maxim (1), it is possible to show that no two distinct theory-contravening candidate revisions of a consistent and logically closed theory by a

sentence α can be set-theoretically compared in terms of the sets of sentences on which they differ with the prior theory.

For the discussion of maxim (2), we need a little bit of terminological preparation. A statement α is *more entrenched* in a theory than another statement β if and only if the agent holds on to α and gives up β when learning that not both α and β are (or may be) true.[20] A new piece of information α is called *moderately surprising* if $\neg\alpha$ is a non-minimal element of the prior theory with respect to epistemic entrenchment. Let us call a revision by α *amnesic* if the revised theory consists of nothing else but $\mathrm{Cn}(\alpha)$ (where Cn is some ordinary well-behaved Tarskian logic); otherwise we call the revision *anamnestic*. Now suppose we want to revise a theory by a sentence α and identify two elements of the initial theory that non-redundantly entail $\neg\alpha$. Then it may well happen, when performing a regular AGM-style revision by α, that the scientist removes the *more* entrenched and retains the *less* entrenched statement. An instance of such a situation can indeed *always* be identified when α is moderately surprising and the revision by α is anamnestic. Thus neither of the two maxims that have always belonged to the core rhetoric accompanying AGM-style theory change models is actually obeyed in these models.

This observation may seem baffling at first sight, but it has turned out to be rather robust. I have argued in Rott (2000) that it does not help to talk about theory contraction in place of theory revision, that it does not help to have recourse to AGM's celebrated representation results that link theory change to some sort of minimisation process, and that it does not help either to restrict attention only to those statements of a theory that are actually true. Yet another way to save the idea of minimal change can be to hold that informational economy should not pertain to theories as sets of statements, but to richer representations of the results of scientific theorising: to dispositions for theory change, or equivalently, to structures that are suitable to guide (iterated) theory revisions. We will address this point in some detail. Before doing this, however, we will take a look at potential norms for theory change.

First, we have to be aware of the fact that the demand for informational economy conflicts with other desiderata. For instance, it competes with the *synchronic* or *static coherence* constraints of *logical consistency* and *closure*. We should give up something when we find ourselves caught in an inconsistency, and we typically have to give up not only the culprit sentence, but also many of the sentences that could be derived with the help of the culprit. This is so, among other things, because we want to maintain the logical closure of our theories. Of course, a change that is minimal subject to the constraints of consistency and closure will in general be bigger than a change that is minimal when no constraints are to be respected. Conservativity is itself a criterion of *diachronic* or *dynamic coherence* (Rott 1999). There are more concepts of coherence that we will have reason to consider below, viz., *dispositional coherence* and *temporal coherence*. At this juncture it is not at all clear whether the latter two concepts

might generate requirements that compete with the requirements of informational economy.

Secondly, it is instructive to contrast the ideas that are advocated in the logical literature on theory revision with ideas recommended by philosophers of science who think about theory change as arising in the (r)evolution of scientific theories and research programmes. In their elementary but thought-provoking little book *The Web of Belief*, Quine and Ullian (1978) list six 'virtues' of hypotheses: empirical adequacy, simplicity, conservatism, modesty, generality, refutability and precision. Logical models of theory change, on the other hand, have appealed to success, logical closure and consistency, informational economy, dispositional coherence, preference and indifference,[21] and temporal coherence (these terms will be explained in due course.) A simple check of these lists reveals that the only criterion that is endorsed both by Quine and Ullian and the standard account of theory revision is informational economy. And a moment's thought makes it clear that at least some of the virtues mentioned compete with one another. Simplicity often requires deviations from informational economy, modesty conflicts with refutability.

Fortunately, we do not have to commit ourselves to one or the other set of criteria here. It is enough at this juncture to note that an account of how to integrate various criteria of *theory choice* (the perspective dominant in the philosophy of science) and *theory change* (the main perspective studied in philosophical logic) is badly needed, but has never been offered. In any case, there is no obvious reason for according informational economy a privileged status among the many contenders that we have identified. It has only a very restricted normative force.

6. CONSERVATISM WITH RESPECT TO STRUCTURES GUIDING THEORY REVISION

We now turn to another interpretation of the idea that scientists should aim at preserving what they have. The propositional content encoded in statements of the scientific object language is not the only kind of valuable information. We may also be interested in preserving the non-propositional information encoded in theory-revision guiding structures, i.e., in richer representations of the results of scientific theorising.

Rich representations of theoretical structures that have gained some currency in the literature include *plausibility orderings* of possible worlds. Given such an ordering, the core set of the most plausible worlds contains those worlds that are consistent with the current scientific theory, i.e., those that could be the real world given what the scientist knows. The ordering of the remaining worlds reflects their relative distance from this core set. Theory change prompted by some new piece of information α then proceeds by manipulating the ordering of worlds in such a way that all the most plausible worlds in the revised ordering

satisfy α. This constraint is mild and leaves a lot of freedom for the exact specification of a coherent revision mechanism. The most *conservative* or *economical* way of changing the plausibility ordering that respects this constraint was defined and investigated by Boutilier (1993, 1996).

Another way of richly representing theoretical structures consists in *entrenchment orderings* of sentences (their interpretation was already mentioned in the previous section). Given an entrenchment ordering, the statements that are least plausible under this ordering are those sentences that are not accepted at all. The ordering of the remaining statements reflects how firmly they are endorsed by the scientist, i.e., the relative tenacity with which the scientist is determined to cling to his statements. In this model, theory change prompted by a new piece of in formation α proceeds by manipulating the ordering of sentences in such a way that α is not among the least entrenched sentences under the revised ordering. Again, this constraint is a rather mild one and leaves a lot of room for filling in the details of an exact specification of the revision mechanism. The most *conservative* or *economical* way of changing the entrenchment ordering that respects this constraint is investigated in Rott (2003a).

It turns out that the account based on entrenchment orderings is essentially a generalisation of the possible worlds account proposed by Boutilier. In the specific context of the assumptions made by Boutilier, the methods are equivalent. Both methods satisfy an axiom for iterated theory change which is sufficient to characterise conservative revisions of richer representations of theories or theoretical structures. As shown by Boutilier (1993, 1996) and, in a more general setting, by Rott (2003a), a repeated conservative change of T first by a statement α and then by a second statement β leads to the same result as a single conservative revision of T by β, provided that β is inconsistent with the result of the revision of T by α. In symbols: $(T*\alpha)*\beta = T*\beta$ if β is inconsistent with $T*\alpha$. Why do we call this recipe "conservative"? Because it suggests that if β cannot be accommodated consistently, the way of handling it in the revised theory $T*\alpha$ is just the same as it was in the original T – as if the agent had never learnt about α. Loosely speaking, the structure of the old theory is stronger than the new piece of evidence α.

Almost immediately after Boutilier had suggested conservative theory change as a natural extension of the AGM model, Darwiche and Pearl (1994, 1997) discovered that the behaviour defined by this model is intuitively queer. They gave the following example. A person who we may for the sake of argument take to be in a state of complete ignorance observes an animal that she takes to be a bird (b). As the animal comes closer, the person perceives that the animal is red (r). A few moments later, she realises (perhaps informed by an expert) that the animal is not a bird after all ($\neg b$). If we use Boutilier's method of conservative theory change, the result of these three subsequent revisions is $\mathrm{Cn}(\neg b)$.

Darwiche and Pearl rightly argue that this is not satisfactory. Why "forget" about the colour of the animal just because it turns out that the person has misclassified the animal as a bird? In more general terms, it can be shown that this

model is *temporally incoherent*. The AGM-postulate of "success" for revisions says – *pace* Neurath – that the most recent piece of information should always be included in the revised theory. Thus, at the moment of receipt the most recent piece of information is treated as the "most important" one. But this privilege is immediately lost when new information happens to come in. To see this, let α, β and γ stand for sentences that are pairwise consistent, but jointly inconsistent. Then iterated conservative change of the trivial theory $\mathrm{Cn}(\emptyset)$ first by α, then by β and finally by γ results in the theory $\mathrm{Cn}(\alpha,\gamma)$.[22] The first and the last piece of information are stronger than the one that comes in between.

Conservatism with respect to revision-guiding structures thus has unacceptable consequences when applied as a method for iterated theory revision. It violates the requirement that a good method of theory change be *temporally coherent*, i.e., coherent in its attitude towards the value of the recency of information. As long as one decides invariably to accept new information (i.e., to take the last piece of information to be the most important one), the only coherent attitude towards the recency of information is to take the second-last piece of information to be the second-most important one, and so on. Instead of $\mathrm{Cn}(\alpha,\gamma)$ as above, the desired result would thus be $\mathrm{Cn}(\beta,\gamma)$. There is in fact an alternative model that yields precisely this result (first studied systematically by Nayak 1994).[23] But as we said before, Neurath's principle that leaves the scientist the free choice whether or not to accept a new sentence (as a protocol statement) is better from a philosophical point of view. Being altogether indifferent to the time at which a certain piece of information is received is a form of temporal coherence, too.

We now leave the field of "economic reasoning" and turn to "economical reasoning", i.e., to the left branch of Figure 1 that refers us to the realm of choices, preferences and utilities.

7. RATIONAL CHOICES AND LOGICAL PROPERTIES

As pointed out above, the classical AGM model of theory revision embodies a strong idea of dispositional coherence. More precisely, the agent's dispositions to change his theory in potential revisions are guided by choices that can be rationalised by a complete preordering, i.e., the agent acts *as if* he was a maximiser with respect to this ordering. All possible worlds are presumed to be comparable with each other in terms of plausibility (AGM 1985, Grove 1988), all statements are presumed to be comparable with each other in terms of entrenchment (Gärdenfors and Makinson 1988). These facts find expression in the seventh and eighth postulates of Alchourrón, Gärdenfors and Makinson (1985) which constrain the scientist's disposition to change his theory:

(*7) $$(T * \alpha) \cap (T * \beta) \subseteq T * (\alpha \vee \beta)$$

which is sometimes called "disjunction in the antecedent", and

(*8) If $\neg\beta \notin T * \alpha$, then $T * \alpha \subseteq T * (\alpha \wedge \beta)$

which is sometimes called "rational monotonicity". These postulates have turned out to correspond to conditions on semantic or syntactic choice functions known as *Sen's properties* α and β^+ which in effect require that the revision function be rationalisable by a complete and transitive preference relation. Many concrete systems of theory revision do not satisfy (*7) and (*8) – a fact that shows that these AGM axioms are very strong. Knowing this, it is pleasing to find that one can draw on the rich resources of the theory of rational choice in order to introduce appropriate weakenings of the theory revision postulates. For instance, a weaker form of (*8) is

(*8c) If $\beta \in T * \alpha$, then $T * \alpha \subseteq T * (\alpha \wedge \beta)$

also called "cumulative monotonicity". This condition corresponds to "Aizerman's axiom" in the theory of rational choice (see Moulin 1985). The theory of rational choice is a powerful instrument which is suitable for analysing and constructing revision operations that are much more flexible than the original AGM ones. Rott (2001, Chapter 7) describes how exactly one can use semantic choice functions (for the selection of the most plausible worlds) and syntactic choice functions (for the selection of the least entrenched sentences) in the construction of theory revisions, and also how postulates for theory revision correspond to rationality requirements for semantic and syntactic choice functions.[24] Seen from this perspective, the theory of theory change can indeed be interpreted as being based on economical principles.

Almost from its beginning, however, the classical theory of rational choice has been subjected to serious criticism. In the rest of this section we present an argument to the effect that a fundamental problem for the theory of rational choice transfers directly to theories of theory revision.

Consider the following example which is not a scientific one in order to keep things simple. A well-known philosophy department has announced a position in metaphysics. Among the applicants for the job there are a few persons we happen to know. First, there is Amanda Anderson, a highly profiled, excellent metaphysician. Second, we have Bernice Becker, who is also definitely very good, though not quite as highly profiled in metaphysics as Andrews. In addition, Becker has done some substantial work in logic. A third applicant is Carlos Cortez. He has a comparatively slim record in metaphysics, but he is widely recognised as one of the most brilliant logicians of his generation.

Suppose that our initial "theory" about the case – i.e., our initial set of beliefs and expectations – includes that neither Anderson nor Becker nor Cortez will get the job (say, because we think that Derek Davidson, an outstanding metaphysi-

cian, is the obvious candidate who is going to get the position anyway). Let us also be clear about the fact that there is only one job available.

Consider now three hypothetical scenarios, each of which describes a *potential* development (*not* a sequence of stages) of the selection procedure. In each of these alternative scenarios we are genuinely taken by surprise, because we learn that one of the candidates we had believed not to stand a chance will be offered the position. To make things shorter, we introduce some abbreviations. Let the letters a, b and c stand for the statements that Anderson, Becker and Cortez, respectively, will be offered the position.

Scenario 1. The dean tells us confidentially that it has been decided that either Anderson or Becker will be appointed.[25] This message boils down to supplying us with the premise $a \vee b$. Given this premise, we conclude that Anderson, being the better metaphysician, will get the job. We also infer that the other candidates will return empty-handed.

Scenario 2. This is a very unexpected scenario in which we are told by the dean that Cortez is actually the only serious candidate left in the competition. Fortunately, there is no need to invest a lot of thinking here. We accept c in this case.

Scenario 3. In this scenario the dean tells us that it has been decided that either Anderson or Becker or Cortez will get the job, thus supplying us with the premise $a \vee b \vee c$. This piece of information triggers off a rather subtle line of reasoning. Knowing that Cortez is a splendid logician, but that he can hardly be regarded as a metaphysician, we realise that competence in logic is considered to be a non-negligible asset by the selection committee. Still we keep on believing that Cortez will not make it, because his credentials in metaphysics are just too weak. Since, however, logic appears to contribute positively to a candidate's profile, we conclude that Becker, and not Anderson, will get the job.

This qualitative description should suffice for our purposes, but for readers who prefer more precision, the following story may help. The selection committee has decided to assign points to evaluate the candidates' work. Anderson scores 97 out of 100 in metaphysics, but has done no logic whatsoever, so she scores 0 here. Becker scores 92 in metaphysics and a respectable 50 in logic. Cortez scores only 40 in metaphysics, but boasts of 99 in logic. In scenario 1, we take it that metaphysics is the only criterion, so clearly Anderson must be the winner. Scenario 2 is trivial. In scenario 3, we gather that, rather unexpectedly, logic matters. As can easily be verified, any weight we attach to logic between $^1/_{10}$ and $^1/_2$ (with metaphysics taking the rest) will see Becker end up in front of both Anderson and Cortez.

Let us now summarise our conclusions from the various premises that the dean supplies us with. Our initial theory T contains $\neg a$, $\neg b$, and $\neg c$ among other things. In scenario 1, the new piece of information $a \vee b$ leads us to accept a and $\neg b$ (along with $\neg c$). In scenario 2, c simply makes us accept that $\neg a$ and $\neg b$. In scenario 3, the new piece of information $a \vee b \vee c$ leads us to accept that $\neg a$ and

b. We can now show that this situation refutes some of the basic logical principles of "economic" theory revision.

First, the example shows that "disjunction in the premises" (*7) does not hold. Take (*7) and substitute $a \vee b$ for α and c for β. Then notice that $\neg b$ is believed if the input is $a \vee b$, and also if the input is c. But $\neg b$ is not believed if the input is $a \vee b \vee c$. Thus the revised theory $T^*(a \vee b \vee c)$ does not include what is common to $T^*(a \vee b)$ and $T^* c$, and (*7) is violated.

Secondly, we find that the situation does not conform to the weakened monotonicity postulate (*8c). Take (*8c) and substitute $a \vee b \vee c$ for α and $a \vee b$ for β. Even though we believe that $a \vee b$ is true if we are given the information $a \vee b \vee c$, it is not the case that everything believed on the basis of the latter is also believed on the basis of $(a \vee b \vee c) \wedge (a \vee b)$ which is equivalent with $a \vee b$. Sentences $\neg a$ and b are counterexamples. Thus the revised theory $T^*(a \vee b \vee c)$ is not a subset of the theory $T^*((a \vee b \vee c) \wedge (a \vee b)) = T^*(a \vee b)$, and (*8c) is violated. *A fortiori*, (*8) is violated as well.

From where are these problems derived? We said that principles of theory change can be systematically interpreted in terms of rational choice. On this interpretation, "disjunction in the premises" (*7) turns out to be an instantiation of one of the most fundamental conditions – perhaps *the* most fundamental condition – of the theory of rational choice: Sen's Property α. This condition, also called *Independence of Irrelevant Alternatives* or *Chernoff property*, says that any element which is optimal in a certain menu remains an optimal element after some other elements have been cancelled from the menu. The three scenarios in our example are modelled after well-known choice situations in which Property α is violated – cases that also happen to disobey Aizerman's axiom. Both properties may fail to be satisfied if the very 'menu' from which the agent is invited to choose carries important information. This phenomenon which Sen calls the 'epistemic value' or the 'epistemic relevance of the menu'[26] suggests that the context of choice has a decisive influence upon the shape of the preference relation of the agent – an idea strongly opposed to the idea of context-independent preferences that underlies the classical theory of rational choice. The *locus classicus* for the problem is a passage in Luce and Raiffa (1957, p. 288) who chose to avoid the problem of the epistemic value of the menu by *fiat*:

> This illustrates the important assumption implicit in axiom 6 [essentially Property α, H.R.], namely, that adding new acts to a decision problem under uncertainty *does not alter one's a priori information as to which is the true state of nature*. In what follows, we shall suppose that this proviso is satisfied. In practice this means that, if a problem is first formulated so that the availability of certain acts influences the plausibility of certain states of nature, then it must be reformulated by redefining the states of nature so that the interaction is eliminated.

This may make good sense as a rejoinder in the context of the general theory of choice and decision. An explanation of how information is surreptitiously conveyed through the particular contents of the menu and how it affects the

chooser's preferences is simply not this theory's business. Unfortunately, the same defence is not available for the problem highlighted by the above job example. It *is* the task of the theory of theory formation to model how one's prior theory is affected by information received from external sources. This is precisely what this theory has been devised to explain, and therefore the anomaly cannot be pushed to a neighbouring research field.

The question raised by our example is a general one. It is hard to get rid of the feeling that the dean's information about the final candidates conveys more information than meets the ear. The fact that a logician gets *mentioned* as a top-ranking contender or that logic becomes a *topic* seems to carry surplus information, over and above the propositional content of the corresponding statement. Does the very fact that a statement is offered in a menu for acceptance have a special relevance for processes of theory revision that has been overlooked so far? Or are there other ways out of the predicament? In any case, the upshot of this section is that the economist's point of view supplies some solid planks for Neurath's boat, but by the same token, it also makes us suspect that these planks cannot be put together without the boat springing a leak.

8. CONCLUSION

This paper started out from Neurath's coherence theory of knowledge and Schlick's claim that Neurath makes implicit use of an economy principle to alleviate the notorious ambiguities inherent in coherentist epistemology. We then reviewed work in the tradition of the AGM approach to theory revision, arguably the most prominent logical paradigm of purifying theories from contradictions. Our overarching questions were to what extent economic(al) principles have played a role in the *actual* development of this paradigm, and to what extent such considerations *should* have been followed. Our conclusions are mostly negative. Informational economy (conservatism with respect to accepted sentences), although widely advertised as *the* central motivation of theory revision models, turns out not to have played anything like a dominant role in the development of such models, and we have found no reason why it should. I think that Schlick was right about that. Conservatism with respect to revision-guiding preferences has in fact been suggested as a strategy for iterated theory revision, but it soon turned out to have unwelcome consequences. So the theory of theory revision has as a matter of fact not focused on *economy*, and the idea of *economical* theory changes has very limited normative force, too. As regards *economic* theory change, our findings are more encouraging. It is possible to reconstruct large parts of theory revision in terms of rational choice theory. So as a matter of fact, ideas coming from *economics* have prevailed in the AGM paradigm and related approaches. However, at the end of the paper we found that a fundamental problem of the general theory of choice seriously infects the specific application area of theory change. Thus the use of rational choice theory has helped us to spot a

new puzzle rather than solve old problems. There is a lot of work in cognitive economics that waits to be done.[27]

NOTES

1. Neurath often emphasised that constructing the set of protocol statements is a matter of choice, e.g.: "Without an adequate selection [of observation statements] at the start, fertile scientific work is impossible." (1931, p. 84)
2. Neurath himself rejected the term "coherence theory" as a label for his point of view, but in his reply to Schlick (Neurath 1934), he does not even begin to comment on Schlick's (re)interpretation of his account as based on the economy principle. One might speculate whether there is a path leading from Richard Avenarius' *Principle of Least Action* or Ernst Mach's *Principle of the Economy of Thought* to Schlick's mentioning of an economy principle here. Mach was, to be sure, one of the most influential scientists at the turn of the century, and Neurath's founding of the *Verein Ernst Mach* in 1928 was one of the major steps in the development of the Vienna Circle. But Mach's principle was much closer in spirit to Occam's razor than to the economy principle formulated by Schlick.
3. The landmark paper is Alchourrón, Gärdenfors and Makinson (1985), for book-length treatments see Gärdenfors (1988), Hansson (1999) and Rott (2001).
4. This insight was called *the Neurath Principle* by Haller (1982a,1982b).
5. See the papers on "non-prioritised belief revision" collected in Hansson (1997).
6. The quotation is taken from McFadden (1999, p. 73). Simon himself famously advocated an alternative model of rationality, viz., that of *satisficing* instead of maximising. According to this model, economic agents set out for achieving a certain level of aspiration, without investing any extra efforts to find out whether they could obtain something even more valuable.
7. According to Rosenberg (1983), economics is not an empirical science, but a branch of applied mathematics – but he does not say applied to *what*. My picture below is more in line with the view that economics is a normative discipline, a theory of rationality.
8. It takes more than transitivity and completeness for a preference relation to be representable by a utility function. An additional technical continuity property is needed (Debreu 1959, 54-59). A concise presentation of ordinal and expected utility theory is given by Hausman and McPherson (1996, Chapter 3: "Rationality").
9. "Egocentric" is not meant to imply "selfish" here. Egocentric agents only look at their personal preference, but the preferences themselves may embody all kinds of (possibly altruistic) thoughts and feelings.
10. Throughout this paper, it is presupposed that *revisions* have to be *successful* in the sense that they efficiently incorporate the specified new piece of information into the current theory. *Contractions* are called *successful* if they efficiently remove some specified statement from the theory (unless that statement happens to be a logical truth).
11. See, for instance, Boutilier (1996, pp. 264-265) and Darwiche and Pearl (1997, p. 2).
12. The first premise does not take into account that one may be forced to give away what is valuable, or that one may invest something at a given time in order to make profit later on.
13. I grant that this train of thought is not very compelling. Alternatively, the second premise might be replaced by another one which, however, is at least as dubious: "It is always better to have something than to have nothing."
14. In the following, a "theory" is meant to be a set of statements that includes its own logical implications. Thus we endorse the logician's idealisation or stipulation that theories be deductively closed.
15. Sometimes use is made of use a partitioning or lumping together of the informational contents of the beliefs .
16. Belief-revision guiding structures typically encode *more* information than the set of currently accepted sentences. They in general allow to retrieve the theory, so that a second component

specifying the belief set is unnecessary. Therefore, it is possible to formally identify the result of scientific theorising with a theory-revision guiding structure.

17. See Alchourrón, Gärdenfors and Makinson (1985).
18. For similar projects in the related field of non-monotonic reasoning, see Schlechta (1996) and Lehmann (2001).
19. In the common numbering of axioms, these are the third and fourth AGM postulates.
20. See Gärdenfors and Makinson (1988) and Rott (2001).
21. Pagnucco and Rott (1999).
22. Both AGM and Boutilier's models assume that the orderings of worlds or sentences involved represents a complete pre-order, i.e., that all worlds and all sentences are comparable as regards their plausibility or entrenchment. So it may perhaps be suspected that the problem of the conservative revision strategy is (at least in part) due to the strong requirements of "dispositional coherence" inherent in the AGM model upon which Boutilier's model is built. However, the unwelcome effects of temporal incoherence remain present in exactly the same way even if *all* of the dispositional requirements of the AGM model are dropped (Rott 2003a).
23. I discuss the merits of this model of "moderate belief revision" in some detail in Rott (2003a). Also compare Nayak, Pagnucco and Peppas (2003).
24. Important earlier work was done by Lindström (1991).
25. We take it for granted in this example that the dean is not playing games with us, that she is not lying, that she has the relevant knowledge etc.
26. Sen (1993, pp. 500-503; 1995, pp. 24-26) has brought the problem to wide attention.
27. Most of the material presented here is discussed in more technical detail in Rott (2000, 2001, 2003a, 2003b). Thanks go to Vincent F. Hendricks, Eckehart Köhler, Hannes Leitgeb, Nils-Eric Sahlin, Gerhard Schurz, Krister Segerberg and Wolfgang Spohn for very helpful comments and discussions.

REFERENCES

Alchourrón, Carlos, Peter Gärdenfors and David Makinson: 1985, 'On the Logic of Theory Change: Partial Meet Contraction Functions and Their Associated Revision Functions', *Journal of Symbolic Logic* 50, 510-530.

Boutilier, Craig: 1993, 'Revision Sequences and Nested Conditionals', in R. Bajcsy (ed.), *IJCAI-93 – Proceedings of the Thirteenth International Joint Conference on Artificial Intelligence*, 519-525.

Boutilier, Craig: 1996, 'Iterated Revision and Minimal Change of Conditional Beliefs', *Journal of Philosophical Logic* 25, 263-305.

Broome, John: 1999, *Ethics Out of Economics*, Cambridge University Press, Cambridge.

Darwiche, Adnan, and Judea Pearl: 1994, 'On the Logic of Iterated Belief Revision', in Ronald Fagin, ed., *TARK '94 – Proceedings of the Fifth Conference on Theoretical Aspects of Reasoning About Knowledge*, Pacific Grove, Cal.: Morgan Kaufmann, 5-23.

Darwiche, Adnan, and Judea Pearl: 1997, 'On the Logic of Iterated Belief Revision', *Artificial Intelligence* 89, 1-29.

Debreu, Gerard: 1959, *Theory of Value*, Wiley, New York.

Edgeworth, Francis Ysidro: 1881, *Mathematical Psychics*, Kegan Paul, London.

Gärdenfors, Peter: 1979, 'Conditionals and Changes of Belief', in Illkka Niiniluoto and Raimo Tuomela (eds.), 'The Logic and Epistemology of Scientific Change', *Acta Philosophica Fennica* 30 (1978), nos. 2-4, 381-404.

Gärdenfors, Peter: 1988, *Knowledge in Flux. Modeling the Dynamics of Epistemic States*, Bradford Books, MIT Press, Cambridge, Mass.

Gärdenfors, Peter, and David Makinson: 1988, 'Revisions of Knowledge Systems Using Epistemic Entrenchment', in Moshe Vardi (ed.), *TARK '88 – Proceedings of the Second Conference on Theoretical Aspects of Reasoning About Knowledge*, Los Altos: Morgan Kaufmann, 83-95.

Grove, Adam: 1988, 'Two Modellings for Theory Change', *Journal of Philosophical Logic* 17, 157-170.

Haller, Rudolf: 1982a, 'New Light on the Vienna Circle', *The Monist* 65, 25-35.

Haller, Rudolf: 1982b, 'Das Neurath-Prinzip – Grundlagen und Folgerungen', in Friedrich Stadler (ed.), *Arbeiterbildung in der Zwischenkriegszeit*, Löcker, Wien, 79-87.

Hansson, Sven Ove: 1999, *A Textbook of Belief Dynamics: Theory Change and Database Updating*, Kluwer, Dordrecht.

Hansson, Sven Ove (ed.): 1997, Special Issue on 'Non-Prioritized Belief Revision', *Theoria* 63, 1-134.

Hausman, Daniel: 1998, 'Economics, Philosophy of', *Routledge Encyclopedia of Philosophy*, electronic edition.

Hausman, Daniel M., and Michael S. McPherson: 1996, *Economic Analysis and Moral Philosophy*, Cambridge University Press, Cambridge.

Lehmann, Daniel: 2001, 'Nonmonotonic Logics and Semantics', *Journal of Logic and Computation* 11, 229-256.

Lindström, Sten: 1991, 'A Semantic Approach to Nonmonotonic Reasoning: Inference Operations and Choice', Uppsala Prints and Preprints in Philosophy, Department of Philosophy, University of Uppsala, 1991:6.

Luce, R. Duncan, and Howard Raiffa: 1957, *Games and Decisions*, Wiley, New York.

McFadden, Daniel: 1999, 'Rationality for Economists?', *Journal of Risk and Uncertainty* 19, 73-105.

Moulin, Hervé: 1985, 'Choice Functions over a Finite Set: A Summary', *Social Choice and Welfare* 2, 147-160.

Nayak, Abhaya: 1994, 'Iterated Belief Change Based on Epistemic Entrenchment', *Erkenntnis* 41, 353-390.

Nayak, Abhaya, Maurice Pagnucco and Pavlos Peppas: 2003, 'Dynamic Belief Revision Operations', *Artificial Intelligence* 146, 193-228.

Neurath, Otto: 1931, 'Soziologie im Physikalismus', *Erkenntnis* 2, 393-431; English translation 'Sociology in the Framework of Physicalism' in O.N., *Philosophical Papers 1913-1946*, eds. Robert S. Cohen and Marie Neurath, Reidel, Dordrecht 1983, pp 58-90.

Neurath, Otto: 1932/33, 'Protokollsätze', *Erkenntnis* 3, 204-214; English translation 'Protocol Statements' in O.N., *Philosophical Papers 1913-1946*, eds. Robert S. Cohen and Marie Neurath, Reidel, Dordrecht 1983, pp 91-99.

Neurath, Otto: 1934, 'Radikaler Physikalismus und "Wirkliche Welt"', *Erkenntnis* 4, 346-362; English translation 'Radical Physicalism and the "Real World"' in O.N., *Philosophical Papers 1913-1946*, eds. Robert S. Cohen and Marie Neurath, Reidel, Dordrecht 1983, pp 100-114.

Pagnucco, Maurice, and Hans Rott: 1999, 'Severe Withdrawal (and Recovery)', *Journal of Philosophical Logic* 28, 501-547. (Corrected reprint in the *JPL* issue of February 2000.)

Quine, Willard V.O., and Joseph S. Ullian: 1978, *The Web of Belief*, second edition, Random House, New York.

Rosenberg, Alexander: 1983, 'If Economics Isn't Science, What Is It?', *Philosophical Forum* 14, 296-314. Reprinted in E.D. Klemke et al. (eds.), *Introductory Readings in the Philosophy of Science*, 3rd edition, Prometheus Books, Amherst 1998, 154-170.

Rosenberg, Alexander: 1995, 'Philosophy of Economics', in Robert Audi (ed.), *Cambridge Dictionary of Philosophy*, Cambridge University Press, Cambridge, 582-583.

Rott, Hans: 1993, 'Belief Contraction in the Context of the General Theory of Rational Choice', *Journal of Symbolic Logic* 58, 1426-1450.

Rott, Hans: 1999, 'Coherence and Conservatism in the Dynamics of Belief. Part I: Finding the Right Framework', *Erkenntnis* 50, 387-412.

Rott, Hans: 2000, 'Two Dogmas of Belief Revision', *Journal of Philosophy* 97, 503-522.

Rott, Hans: 2001, *Change, Choice and Inference*, Oxford University Press.

Rott, Hans: 2003a, 'Coherence and Conservatism in the Dynamics of Belief. Part II: Iterated Belief Change Without Dispositional Coherence', *Journal of Logic and Computation* 13, 111-145.

Rott, Hans: 2003b, 'A Counterexample to Six Fundamental Principles of Belief Formation', manuscript, January 2003.

Schlechta, Karl: 1996, 'Some Completeness Results for Stoppered and Ranked Classical Preferential Models', *Journal of Logic and Computation* 6, 599-622.

Schlick, Moritz: 1934: 'Über das Fundament der Erkenntnis' in *Erkenntnis* 4, 79-99; English translation 'The Foundation of Knowledge', in Alfred J. Ayer (ed.), *Logical Positivism*, New York, Free Press 1959, 209-227.

Sen, Amartya K.: 1993, 'Internal Consistency of Choice', *Econometrica* 61, 495-521.

Sen, Amartya K.: 1995, 'Is the Idea of Purely Internal Consistency of Choice Bizarre?', in J. E. J. Altham and R. Harrison (eds.), *World, Mind, and Ethics. Essays on the Ethical Philosophy of Bernard Williams*, Cambridge University Press, Cambridge, 19-31.

Spohn, Wolfgang: 2003, 'Die Logik und das Induktionsproblem', in Peter Schroeder-Heister und Wolfgang Spohn (eds.), *Logik in der Philosophie*, Synchron-Verlag, Heidelberg.

Department of Philosophy
University of Regensburg
D-93040 Regensburg
Germany
hans.rott@psk.uni-regensburg.de

NILS-ERIC SAHLIN

INFORMATIONAL ECONOMY AND CREATIVITY

Comments on Hans Rott's "Supplying Planks for Neurath's Boat"

We all know that Ramsey argued that the best way to understand how the theoretical entities of a theory function is to picture them as existentially bound variables. If the entities of our theory are α, β, and γ, the best way to write our theory, he tells us, is: $\exists\alpha,\beta,\gamma$ (dictionary & axioms). The theory's 'Ramsey sentence', the existentially bound variables are the carriers of ontological commitment. If the Ramsey sentence is true, they tell us what there is.

Ramsey sentences have been used trying to eliminate so-called theoretical terms in favour of so-called observational terms. Ramsey's own aim was, however, not to do away with the theoretical terms. In fact, he had an argument showing that, depending on the type of dictionary used (containing definitions connecting theoretical expressions and observational ones), and the type of definitions introduced (for example explicit definitions), such a strategy leads to static theories. Ramsey's goal was to explain the function of theoretical terms. It is worth noting that to do this, he does not use the dictionary to 'define' the theoretical terms in terms of the observational terms. Instead, he does the opposite. He uses the dictionary to define the observational terms of the 'primary language' (the observation language) in terms of the theoretical terms of the 'secondary language' (the theoretical language). This gives us a profound understanding of how the two type of terms work together in a theory.

Ramsey's view of theories has several advantages. First, Ramsey sentences help us understand the dynamics of scientific theories and scientific growth. Second, they explain the phenomenon of 'incommensurability.' We note, for example, as Ramsey explains, that no proposition of a theory

> can be understood apart from the whole theory to which it belongs. If a man says "Zeus hurls thunderbolts", that is not nonsense because Zeus does not appear in my theory, and is not definable in terms of my theory. I have to consider it as part of a theory and attend to its consequences, e.g. that sacrifices will bring the thunderbolts to an end.

Thus, the 'adherents of two such theories could quite well dispute, although neither affirmed anything the other denied'. And something similar holds for terms like 'belief' and 'minimal revision.' As Ramsey explains, any additions to our theory, be it a new axiom, a "particular assertion" or a new definition, have to be made within the scope of a theory's quantifiers. And making additions, we have to consider what else we want to add in the future, and whether the addition or its

F. Stadler (ed.), Induction and Deduction in the Sciences, 247–249.
© 2004 *Kluwer Academic Publishers.*

negation is the one best suited for upcoming revisions. We can dispute between theories and how to expand them, but our reasoning, taking place within a theory, within the scope of the quantifiers, is not affected. Ramsey was an ontological anti-realist – theoretical terms acquire their meaning by their function in the theory.

But Ramsey's analysis also tells us something about another important subject: creativity. The problem-solver is working "within" the Ramsey sentence. The reasoning takes place within a theory, within the scope of the theory's quantifiers – he or she is deducing theorems, testing consequences, and all that. The creative problem solver, on the other hand, works on the theory – dispute between theories. Looking at the Ramsey sentences, we notice that in fact there are only two ways in which we can be creative. We can either add new concepts to the theory or we can change its axioms and dictionary, i.e., its rules. There are two fundamental forms of creativity: concept creativity and rule creativity. In addition, the Ramsey sentences tell us that these two types of creativity are the two sides of one and the same coin. If we, for example, are concept creative, if we add new things to the Ramsey sentence, we are forced to change its rules, i.e., the axioms and dictionary. And if we, in a non-trivial way, change the axioms of the theory, it will affect our concepts.

Hans Rott stresses that the criterion of informational economy may not have played that important role for the development of such theories as, for example, the AGM theory or Levi's theory of belief dynamics. But Gärdenfors and Levi are discussing and advocating various principles of minimal change: expansion, revision and contraction. Some time ago, Peter Gärdenfors, Bengt Hansson and I met frequently to discuss these types of ideas and problems and later we met to go through the *Knowledge in Flux* manuscript. We asked questions like "why *minimal* revisions?," "how do we measure revisions?", "do we need a measure of entrenchment?", "how are belief revision theories related to decision theories and theories of evidence?" Questions of informational economy were definitely on the agenda. But that is history. Today my complaint is that however interesting these theories are, they tell us little if anything about creativity – they are in one fundamental respect static theories. One might argue that they were never designed to mimic creativity, to analyse and portray truly dynamic belief change and concept formation, so my criticism is unfair. That is true. But as important as it is to look at "theory change performed economically", it is also important to look at creative theory change – how new concepts and rules helps us solve problems that cannot be solved by moving around within a "fixed" belief structure. That is help for us in moving from one Ramsey sentence to another. Neither the AGM theory, nor Levi's theory, can tell us how to do this. And the question is if that type of change can be performed economically. In fact, belief changes performed economically seem to be self-defeating, having to satisfy various principles of minimal change means that one overlooks the best solution to the core problem.

It is important to note that this shortcoming of the theories has nothing to do with the choice of logic. The criticism is valid only if we are working within the framework of classical logic. That we, with nonmonotonic reasoning, can get the creativity we are looking for. That is not true. A well-known theorem of Gärdenfors and Makinson tells us that: if we assume that there is an ordering of our expectations, which satisfies three rather natural conditions, transitivity, dominance and conjunctiveness, then nonmonotonic logic is nothing but classical logic with relevant expectations added as explicit premises. Indirectly, this theorem tells us something very important. Creativity has very little to do with inferences, with the reasoning taking place within the scope of the quantifiers. Nonmonotonic inferences are as uncreative as monotonic inferences.

REFERENCES

See Hans Rott's paper and:

F.P. Ramsey, *Philosophical Papers*, ed. by D.H. Mellor, Cambridge University Press, Cambridge 1990.

N.-E. Sahlin, *The Philosophy of F.P. Ramsey*, Cambridge University Press, Cambridge 1990.

N.-E. Sahlin, *Kreativitetens filosofi*, Nya Doxa, Stockholm 2001.

Department of Philosophy
Lund University
Kungshuset, 222 22 Lund
Sweden
nils-eric.sahlin@fil.lu.se

I. GRATTAN-GUINNESS

THE PLACE OF THE NOTION OF CORROBORATION IN KARL POPPER'S PHILOSOPHY OF SCIENCE

1. CORROBORATION IN POPPER'S PHILOSOPHY

The main text of Popper's *Logik der Forschung* (1935) consists of seven chapters outlining the main features of his falsificationist philosophy of science, followed by two involving probability theory and quantum mechanics, and finally one on corroboration, which is the main concern here. (I shall cite his English translation of the book (Popper 1959), noting its newer material when appropriate.) The chapter begins with a section on non-verifiability of theories, which would have been better placed in chapter 6 on testability, and then two sections on the probability of events and of hypotheses and on probability logic, which belong rather in chapter 8. Only then come four short sections on corroboration, the property of a theory that it passes its tests. The chapter is placed last because of a link between corroboration and probability; he even claimed that there was no difference between the two in that both are appraisals (1959, 265). In a later appendix *ix, and in more detail in other work (1982a, 240-255), he offered quantitative measures of corroboration involving probabilities, but at the end of the latter text he almost dismissed the enterprise; and in similar vein, I shall not consider it further, since the measures depend upon the tests which happen to have been carried out and so are not very significant. Otherwise, he did not discuss corroboration very much in his writings, and it is not well handled by commentators;[1] but this rather low status contrasts markedly with the actual practice of science, where scientists and engineers regularly work with (seemingly) highly corroborated theories, even to the extent that 'reliability science' is a subject in its own right.

I shall argue that corroboration deserves more attention in falsificationist discussions than it often receives, and that interesting philosophical and educational questions arise beyond its supposed relationship to (or contrast with) verification in scientific induction. I shall use Newtonian mechanics as my main example, as it is an important theory whose main features and applications should be reasonably familiar to the reader. Thus, there is a skew towards the physical sciences and technology; but I am sure that the points made apply also to the natural and human sciences.

F. Stadler (ed.), Induction and Deduction in the Sciences, 251–262.
© 2004 *Kluwer Academic Publishers.*

2. THE CONTENT AND STATUS OF CORROBORATION

The basic tenets of corroboration held by Popper are these. It is a qualitative measure of the success that a theory has achieved in passing tests; that is, that the predictions made from it are borne out by experience or experiment, at least as well can be determined under the conditions and precision of the tests and experimenters and observers. It has no cumulative or inductive character at all; repeated success has no bearing whatsoever on the outcome of the next test, nor is falsifiability compromised. Thus, it differs fundamentally from verification as handled in inductivist philosophies of science;[2] indeed, when considerations of the probability calculus are brought to bear upon corroborability, *im*probability increases. It is a methodological notion, to be distinguished from the verisimilitude of a theory, which is a semantic notion that records the extent to which its predictions do (not) correspond to the pertinent facts.[3] As with every other procedure envisioned in this philosophy, including falsification, corroboration can be fallible.

In a later footnote to his book, Popper suggested that for a theory T which enjoys a range of domains of application, corroboration be construed not as a property of T but as a relation between T and a domain within this range (1959, 412). This was a good modification: to take an example to which I shall soon return, Newtonian mechanics is far more highly corroborated in celestial mechanics than in engineering mechanics. Sadly, he rejected his modification at once, but I shall adopt it when it is useful, not only for corroboration, but also for testing in general: a theory is (or is not) corroborated in some domain of application. I also adopt the distinction stressed in (Bogen and Woodward 1988) between data needed to provide inferences about the phenomena at hand, and the phenomena themselves from which falsifying or corroborating evidence for the theory is drawn. I take the discussion below to apply to theories descriptive of the phenomena under study and those which attempt to offers explanations of them.

The low status of corroborations in falsificationism is well exemplified by a philosophical problem that was very influential on young Popper in the formation of his philosophy. In 1919 Newtonian mechanics was found to provide incorrect calculations of the advance of the perihelion of the planet Mercury and of the refraction of light passing the edge of the Sun, and much of the discrepancies between prediction and calculation from observations were found by using general relativity theory, which thereby was given a great boost (Jeffreys 1919). But the following feature of the reaction is rarely noted. There was much concern with this discrepancy *only because Newtonian mechanics had already achieved a massive catalogue of corroborations in celestial mechanics; moreover, the two falsifications were irrelevant to almost all of them.* Had this situation not obtained, the discrepancies would have aroused little interest. This point applies

even if relativity theory had not had been corroborated by the new data, though the corroboration added greatly to its prestige.

Despite this setback, Newtonian mechanics was and is still very widely used. This shows that *we can still use falsified theories in domains of application away from those where falsification has happened* (Popper 1959, 53-54), in this case in terrestrial and planetary mechanics and even in various other parts of celestial mechanics.

Corroboration also plays an important role in Popper's claimed solution to 'the problem of induction' (1972, ch. 1). In fact, he claims to solve two problems: the epistemological one, where he asserted that no guarantee exists of invariance or regularities from the past to the future; and a methodological problem, where he advised that if competing theories are available for a choice of a course of action we should choose the 'best-tested theory'. But this is a serious slip of the pen; the context makes clear that he meant the best *corroborated* theory. However, even that version of his advice meets difficulties, for critics have pointed out that theories which propose an ad hoc modification of the original theory T at some give future moment of time N are just as well corroborated (and tested) up to N as is T itself, and Popper had to resort to special metaphysical pleading for theories to be invariant over space and time (Worrall 1989). However, the logical analysis of the propositions involved in testing shows that an alternative logical layout of them can be proffered in which such special pleading is avoided (Grattan-Guinness 2003).

3. THE STATUS OF NORMAL SCIENCE

Falsification corresponds to the testing *to destruction* of a theory. Now this is very important, just as in the way that, say, a car manufacturer should test his models to destruction in order to assess their robustness and behaviour in accidents. But I buy a car to fulfil much less extreme tasks, and hopefully always safely. In the same way, successful scientific theories gain most of their importance from the corroborations that they (seem to have) achieved, not from their ultimate destruction. Like many other philosophers, Popper tended to focus upon the 'great moments' of science such as that of 1919, when falsification stared a major theory in the eye and a competitor stood waiting in the wings; but this focus is rather naïve, resembling the approach of historians who emphasize only Great Events and the Great Men involved in them. Most of the time in science, philosophy and history, far less crucial processes are taking place; but they may well involve major theoretical and/or experimental difficulties.

These remarks are congenial with Thomas Kuhn's characterization of normal science (Kuhn 1962): paradigms (or several of the many types of them which Kuhn proposed) set the style of theorizing for many ranges of application of the science concerned, and corroboration is normally expected. When some major 'revolution' (or maybe an innovation) takes place, then patterns of basic thinking

in this context have to shift. While finding Kuhn's account *least* useful in explaining the structure of revolutionary changes, I accept his advocacy of the *Gestalt* nature of the change: to continue with the above example, after relativity theory the relationship between space-time and velocity and acceleration had to change, and one works either in the old Newtonian mode or in the new one but not with the two together.

The warnings given in (Popper 1970) of the 'dangers' of normal science are well taken, in particular the degeneration of science into mere puzzle-solving, to use Kuhn's particularly unfortunate word (1963, ch. 2).[4] Other dangers include scientists' elevation of a favourite theory into a quasi-religion in which belief gains an excessive status, and the complacent dismissal of data as 'anomalous' when they are naughty enough to falsify it. Popper's position was reinforced by his view of science as the study of nature, *not* of consensus among the experts about its character, a view that Kuhn also affirmed.

On the other hand, Popper's attack on normal science once again demoted corroborations, and the positive sides of Kuhn's conception are greatly underestimated. Those who dismiss normal science as robotic exegeses of given principles are profoundly mistaken (they might try to learn some). In cases familiar to me it is hard work in all respects, creative and experimental. Notions and theories subordinate to the main theory can still require deep new notions, leading maybe to new sub-theories normal within their domains, and so on further (for example, from Newtonian mechanics to hydrodynamics to the same for viscous fluids to the same at high temperatures).

4. SCIENCE, TECHNOLOGY AND RELIABILITY

The special bearing of corroboration upon technology (including engineering) needs to be emphasized. I make no distinction between theories and tests being applied to natural phenomena or to man-made materials and artefacts. A theory in chemistry is under test whether it involves, say, pure common salt or some artificial compound; likewise, mechanics can be applied equally to the motions of Mercury or to the motions of a pendulum clock.

The main bearing upon corroborations is as follows. In Popper's view, science is a risk-taking enterprise where theories are formed and then tested as severely as possible. Now science and technology have a very close relationship, and yet technology requires reliability in the performance of its product. *Thus science is risk but technology is safety* – a "paradox" of which the resolution requires careful attention to be paid to corroborations.

An important concern in technology is determining the values of parameters involved in the specification of materials and the design of artefacts; typically, the breaking strength of the types of steel used in a building, and the consequent efforts to ensure that it is not exceeded. Decisions taken here may well be affected by further factors, such as cost (Rescher 1978, esp. pt. D).

Reliability and safety in technology are sufficiently important in technology to have given rise to 'reliability theory'. This subject treats the properties and (in)frequencies of failure of materials, equipment and artefacts, and the strategies for their maintenance and replacement. Largely an interaction between probability with statistics and technology, some of its features were heralded by the introduction in the 1930s of quality production control (see especially Shewhart 1931). It was launched in a large scale after the Second World War, being inspired by many technological novelties and production problems that had been faced then (Barlow 1984). It makes much use of statistical distributions, including some little-known elsewhere; in particular the Weibull, which was originally formed to express the breaking strength of materials but has since been used in many other reliability contexts.

Reliability theory has become a quite wide-ranging subject, taught in many technology courses.[5] However, it seems to be presented as a collection of techniques, without reference to philosophical consequences; and conversely, it seems to be little known to philosophers of science. Yet the distinction between the two concerns is only one of practise, not of knowledge as such: the failure of a light bulb at the 737th switch-on is just as much a falsification of its theory of action as is the motion of Mercury a falsification of Newtonian mechanics. Further, the successes of the previous 736 switch-ons are corroborations of its theory of action just as much as those enjoyed by Newtonian principles elsewhere in mechanics. Other points of contact with the philosophy of science include the use in both subjects (though not usually by falsificationists) of Bayesian techniques; and the reliability expected (but maybe not delivered) of instruments and machines used in testing scientific theories.

The role of probability in reliability theory seems to be philosophically neutral – hence the lack of philosophical discussion in its literature. The distributions are usually applied to assessing repetitions of action or behaviour, sometimes using large runs of data; thus the most natural philosophical interpretation of its theories is frequentist. But its theorems are equally intelligible to, say, a Popperian advocate of probabilities as propensities. Reliability, strong corroborations (up to now, anyway); none of these features of theories in science or technology lies outside the philosophical remit (Ziman 1978).

Reliability theory also takes due note of unreliability: failures in technology, which involve falsifications of theories. These failures are of different kinds: among others, deliberate failure, as in built-in obsolescence of equipment; vulnerability in design, as in the quick collapse of the towers of the World Trade Center in New York after an attack which imposed conditions on the buildings far beyond their designed limits; and faults in preparation of materials, as that of the rivets in the sinking of the *Titanic*. The task of determining the cause of failure can be complicated and even controversial, as in science.

5. THE PHILOSOPHY OF SMALL EFFECTS

The case of the perihelion of Mercury involves another feature of testing theories which can occur in quantified scientific theories of all kinds and which makes the contrast between corroboration and falsification less straightforward than it appears: the discrepancy was an *extremely small quantity*. Two features are worth noting.

Firstly, as mentioned earlier Newtonian mechanics was already highly corroborated, so that the discrepancies did not *obviously* raise grave issues; maybe some modified details (of perturbation theory for Mercury, say) could resolve the issues. In that test, examining the 'next decimal place' (Richtmyer 1932) led to major consequences; but usually this is not the case, and it hard to see how a scientist could tell in advance whether or not a major difficulty is at hand.

Secondly, while in this case the quantity involved did corroborate one theory and falsify another one,[6] in many cases the magnitude of the discrepancy might fall within the range of experimental, observational and human imprecision pertaining to the test, thus preventing any clear conclusion to be drawn. Imagine a situation in which the precision of observation and experiment is only around 10^{-3} in some unit of relevant measurement, and the discrepancy between a prediction by T and the experimental data is a little larger than 10^{-3}. Has T been refuted or not? It is hard to say, and the usual strategy will be to collect more data – while recognizing that *exact* repetition cannot be achieved (Popper 1950) – and maybe even effect some statistical assessment of range of data obtained. In situations where the theory under test is itself statistical, Popper advocated assessment involving likelihood (1959, 410-19).

6. THE PLACE OF TRUTH IN POPPER'S PHILOSOPHY

An important event for Popper was his meeting with Alfred Tarski in 1934 when he learned of Tarski's semantic correspondence theory of truth, which took the truthhood of T away from the perplexing epistemologies of belief and assertion and placed it instead in the safe haven of the meta-theory of T. The distinction between object-theory and meta-theory became and remained central to Popper's work; indeed, some of his best contributions are meta-philosophical, such as arguing for indeterminism and against determinism (Popper 1982b).

Popper adopted this theory of truth in harness with his emphasis upon the truth-*content* of theories. However, in thrusting it to the center of his concerns, he gave it an epistemological weight that Tarski always eschewed (Grattan-Guinness 1984, with special reference to Popper 1972, ch. 9). Popper seems to have thought it to be necessary and sufficient for his purposes, but in fact further assumptions are required for sufficiency to be achieved. The assignment of cor-

respondence between fact and proposition is itself a convention. Again, the restriction of truth to bivalency somewhat restricts the applicability; pace Tarski's own example 'snow is white', perception of colours is hardly a yes-or-no matter, and bivalency imposes a discontinuity upon a continuous spectrum.

More seriously, Popper's emphasis on the facts to which true propositions correspond sounds like empiricism on the rise: theories are appraised *merely* on the resemblance of their predictions to the pertinent facts, without concern for their contents, which may differ radically (such as in our example of Newtonian mechanics versus relativity theory). This restriction does not square with Popper's important point that scientists *think up* theories, which may invoke notions which transcend experience. (Ever felt a Newtonian central force?) It also underrates the unavoidable role of theories both in the expression of facts themselves and in the decision to regard them as relevant to the theory in the first place. Again, it mitigates against his advocacy of common-sense realism but rejection of common-sense epistemology (Popper 1972, ch. 2). It is consistent with philosophies of science to which Popper does not subscribe, such as proposing that science in general is progressive, and perhaps inevitably so.[7]

In my view, there is *no* contradiction in the situation where a theory T_1 has all the corroborations of T_2 and more of them and yet is also further away from the actual state of affairs (whatever they are). Suppose, for example, that light actually is a waval phenomenon: then particulate theories are more distant from actuality than waval ones. Yet before the arrival of Fresnel in the mid-1810s, particulate theories had provided the more elaborate and testable theories in optics for a century.

Popper's scheme (1972, esp. pp. 242-244)

Problem — tentative theory — error elimination — new problem

seems vulnerable to the same doubts. Error is to be eliminated at the level of correction of falsifications by new theories that (hopefully) gain fresh corroborations; but the theoretical structure and assumptions of the old and new theories are not considered. Part of the difficulty here arises from Popper's regrettable silence, in all of his philosophical writings on science, over the *formation of scientific problems in the first place* – in particular, that it is subject to meta-problematic criticism. In particular, it led to him to an overly narrow conception of the role of experiments and observations in science as being subservient to theory in merely providing the means of determining corroboration or falsification (1959, 106-111). In real scientific practice, however, experimental work may set up the problem in the first place, and the relationship with theory becomes much more dynamic (Franklin 1986). For example, why did a certain experiment on an acid not work when the temperature of the acid was very low? Does this failure also obtain with other acids? In such a way a problem is created.

A preferable approach distances Popper's falsificationism from the empiricist elements just noted. It preserves Tarski's correspondence theory of truth, but

starts from Tarski's own appraisal of it rather than Popper's – that it requires a realist ontology but otherwise is *neutral* over epistemological assumptions.[8] Truth and falsity are confined to the factual record, of passing or failing tests. Theories as such are held to be 'ontologically (in)correct' when they do (not) correctly describe the way (whatever it is) in which the phenomena actually happen (Grattan-Guinness 1986). Of course, we shall never know if a theory is correct, because the next test could lead to a falsification; this is the mate of Popper's distinction between truth, possession of truth, and proof of possession of truth.

Concerning the formation of scientific theories, I lay emphasis upon *reification.* The scientist is often faced with a vastly complicated situation S in which the phenomena are taking place; so in order to set up a soluble problem (Medawar 1967), he simplifies S to s and treats it *as if,* for example, the environment is at constant temperature, the Earth is not rotating, the phenomena are taking place in a vacuum, and so on – factors which hopefully have only small effects upon S in the sense discussed above.[9] Then (let us hope) he formulates a theory T about testable s, which is falsifiable at least within the current apparent limits of experimental accuracy. If success follows, in the form of corroborations and/or resolvable falsifications of T, then he puts back some of these factors in a sequence of *desimplifications*: allow first for, say, air resistance (somehow), then the rotation of the Earth, then maybe its precession and nutation, then.... Thereby, s gradually returns towards S via successively more developed versions of T, with the simplifications reduced though probably not eliminated.[10] This sequence of theories could be *locally* progressive in the sense of achieving comparably higher measures or precisions of corroboration, at least in some domains(s) of application; but a *general* progressionist philosophy of science is not entailed. Indeed, different desimplifications may work better in different domains of application of T.

Another benefit of working with reification is the role needed for Popper's 'world 3' of abstract objects such as problems, theories as such, mistakes, and so on (1972, esp. ch. 4; 1994, chs. 2-3). The notion itself is quite legitimate – Popper was not the first to advocate it – but his version has been justly criticized (for example, in Cohen 1980). The doors of world 3 seem to be too wide open, for anything can enter it. One reason for this liberality may be the construal of truth (and also verisimilitude) to the predictions of theories without regard for their respective contents. But when reification is admitted, then one grants membership to notions (of all kinds) only when they are required by a problem, theory or test; then its population can be reined in to some extent.

7. SOME REMARKS ON SCIENCE EDUCATION

In order to be a scientist, almost certainly education will be required. As is well known, science education can be a very off-putting experience to the student, not

only or necessarily because of the inherent theoretical and/or experimental difficulties, but also because of the theological way in which scientific theories are presented: corroboration raised to the level of unquestioned truth, normal science becoming factory-like science, science portrayed as a non-human activity. This manner of presentation seems particularly regrettable but partly unavoidable at school level, when the science conveyed is inevitably rather elementary, and therefore less likely to be falsifiable. There is a great problem of balancing the need for skills with the excitement of original creation, even at a modest level (Ravetz 1972). Popper's philosophy may sometimes be mentioned, but the normal reaction to the failure of a student to corroborate a theoretical prediction is to claim that he made a mistake somewhere – and indeed the claim is almost certainly correct!

The philosophy of science education is a vast topic, and I make only two points here. Firstly, the notion of corroboration should be emphasised, and distinguished from verification: at least some basic points are conveyable, especially that theories can be well corroborated in at least some domains of application, and so are well worth learning! Secondly, there have been quite strong initiatives in many countries to import the history of science into teaching, where the difficulties and even blunders of the past can be conveyed sympathetically along with the corroborations, and the human element introduced (Matthews 1994).

Now falsificationism (with Popper's extension, critical rationalism) possesses a nice point of contact here, which is denied to many other philosophies. For in that philosophy a theory exists at all because it is the outcome of treatment, especially falsifications, of previous theories. (Even in a completely new field of study at least analogies from established theories were probably used, fallibly of course, as sources of inspiration for its theories.) Therefore, falsificationism and history fit together well, although the student does not need to learn all the historical details and complications.[11] Popper himself was interested in the history of science to a degree most unusual in a philosopher of his (or our) time; and while this was in part a personal inclination, it was surely also inspired by its natural place in the kind of philosophy which he advocated.

8. CONCLUDING REMARK

Some major philosophies rest upon platitudes, which the inaugurating philosophers realized contained profound and general insights. Tarski's semantic theory of truth, for example, based upon the distinction between a language and its metalanguage, is grounded upon the platitude that we have to stop our work sometimes and reflect upon its purpose and content. (The logician Kurt Gödel is also to be credited with this insight, which was also partly grasped by some predecessors.) Popper's philosophy is based upon the platitude that we learn from our mistakes, an assertion of naïve fallibilism that he built into falsifica-

tionism and then on to (comprehensively) critical rationalism. The theme of this paper hangs upon another platitude: that we learn from our successes.

NOTES

1. For example, the four articles in (Schilpp 1974) which treat corroboration seem to be among the weakest in the book; Popper's replies are often unavoidably dismissive (1974, 993-999, 1037-1048).
2. Most inductive epistemologies have been cumulative; but some are eliminative, in claiming that when all other theories have failed, the one left must be true. An interesting foray of this kind is (Nicod 1924), who in its course grasped some features of the logic of falsificationism (see esp. ch. 2); however, his intention was quite different from that of Popper, for whom, of course, all theories could be false.
3. Popper once wondered if greater corroboration of a theory may go along with its increased verisimilitude (1963, 234); but he never intended the former to *measure* the latter (1974, 1011).
   I note without discussion the well-known difficulties about the definition of verisimilitude, as my main concern lies in its basic difference from corroboration For discussion, see Fetzer (1981, esp. pt. 3), Dilworth (1981, esp. ch. 3) and Oddie (1986); and for a review of the more recent theories, see Kuipers (1987). The distinction is often missed in other philosophies of science.
4. The misfortune arises from the fact that puzzles, as normally understood, are exercises within a *finite,* though maybe large, space of possibilities, and solutions may be found by exhaustive inspection. But rarely are even the routine uses of science finite in this sense.
5. For a good survey of the main statistical techniques deployed, see Ebeling (1997). The influence of Shewhart (1931) has been substantial in its own right, especially for its role in the ideas on production reliability and management promoted after World War II by W. Edwards Deming.
6. I do not discuss here the non-trivial 'Duhem-Quine' thesis about the under-determination of theories by observations.
7. For discussions of cumulation see Bachelard (1928, esp. bk. 3); Niiniluoto (1984, esp. chs. 4-6); and Dilworth (1981, esp. ch. 5).
8. Thus the confirmationist philosopher Carnap was just as impressed with Tarski's theory as was Popper (for historical context see Wolenski and Köhler 1999, esp. pts. 1-3).
9. There is no necessary link between this 'as if' philosophy and that advocated by Hans Vaihinger in the 1910s, especially not his relativism. However, exploration of similarities may be worth undertaking.
10. An outstanding example of successive desimplifications is the history of the so-called 'simple pendulum', which is anything but simple when many decimal places are required from the observations. Numerous small(ish) factors were considered: not only air resistance and rotation of the Earth but also flexure and torsion of the suspending wire, the non-planarity of the swing, possible differences in time of the up- and down-swings, possible angular momentum at the end of each swing, flexure of the supports of a large pendulum, ... (Wolf 1889-1891).
11. In advocating roles for history in mathematics education, I advocate the method of 'history-satire', where the broad historical record is respected but most of the barnacles are avoided or only treated lightly (Grattan-Guinness 1973). The same kind of procedure could work for the various sciences, though maybe in less detail than is possible for mathematics.

REFERENCES

Bachelard, G. 1928. *Essai sur la connaissance approchée,* Paris: Vrin. [Several reprints.]

Barlow, R.E. 1984. 'Mathematical theory of reliability: a historical perspective', *IEEE Transactions on reliability, R-33,* 16-20. [Followed on pp. 21-133 by articles on the teaching of reliability science, and on its development in various areas of technology.]

Bogen, J. and Woodward, W. 1988. 'Saving the phenomena', *Philosophical review, 97,* 303-352.

Dilworth, C. 1981. *Scientific progress,* Dordrecht: Reidel.

Cohen, L.J. 1980. 'Some comments on third-world epistemology', *British journal in the philosophy of science, 31,* 175-180.

Ebeling, C.E. 1997. *An introduction to reliability and maintainability engineering,* New York: McGraw Hill.

Fetzer, J. 1981. *Scientific knowledge. Causation, explanation, and corroboration,* Dordrecht: Reidel.

Franklin, A. 1986. *The neglect of experiment,* Cambridge: Cambridge University Press.

Grattan-Guinness, I. 1973. 'Not from nowhere. History and philosophy behind mathematical education', *International journal of mathematical education in science and technology, 4,* 421-453.

Grattan-Guinness, I. 1984. 'On Popper's use of Tarski's theory of truth', *Philosophia, 14,* 129-135.

Grattan-Guinness, I. 1986. 'What do theories talk about? A critique of Popperian fallibilism, with especial reference to ontology', *Fundamenta scientiae, 7,* 177-221.

Grattan-Guinness, I. 2003. 'The logic of scientific testing: Did Karl Popper solve "the principle of induction"?', *Erkenntnis,* to appear.

Jeffreys, H. 1919. 'On the crucial tests of Einstein's theory of gravitation', *Monthly notices of the Royal Astronomical Society, 30,* 138-154. Also in *Collected papers,* vol. 6, London: Gordon and Breach, 1977, 176-192.

Kuhn, T.S. 1962. *The structure of scientific revolutions,* Chicago: University of Chicago Press). [2nd ed. 1972.]

Kuipers, T. van 1987. (Ed.) *What is closer-to-the-truth?,* Amsterdam: Rodopi.

Matthews, M.R. 1994. *Science teaching. The role of history and philosophy of science,* New York and London: Routledge.

Medawar, P.B. 1967. *The art of the soluble,* London: Methuen. [Paperback reprints.]

Nicod, J. 1924. *Le problème logique de l'induction,* Paris: Alcan. [English transs. in *Foundations of geometry and induction* (1930) and *Geometry and induction* (1969), London: Allen and Unwin.]

Niiniluoto, I. 1984. *Is science progressive?,* Dordrecht: Reidel.

Oddie, G. 1986. *Likeness to truth,* Dordrecht: Reidel.

Popper, K.R. 1950. 'Indeterminism in quantum physics and in classical physics', *British journal in the philosophy of science, 1,* 117-133, 173-195.

Popper, K.R. 1959. *The logic of scientific discovery,* London: Hutchinson.

Popper, K.R. 1970. 'Normal science and its dangers', in *Criticism and the growth of knowledge* (ed. I. Lakatos and A. Musgrave), Cambridge: Cambridge University Press, 51-58.

Popper, K.R. 1963. *Conjectures and refutations,* London: Routledge and Kegan Paul.

Popper, K.R. 1972. *Objective knowledge,* Oxford: Clarendon Press.

Popper, K.R. 1974. 'Replies to my critics', in P.A. Schilpp (ed.), *The philosophy of Karl Popper,* 2 vols., La Salle, Illinois: Open Court, 961-1197.

Popper, K.R. 1982a. *Realism and the aim of science,* London: Hutchinson.

Popper, K.R. 1982b. *The open universe. An argument for indeterminism,* London: Hutchinson.

Popper, K.R. 1994. *Knowledge and the body-mind problem* (ed. M.A. Notturno), London: Routledge.

Ravetz, J.R. 1972. *Scientific knowledge and its social problems,* Oxford: Clarendon Press.

Rescher, N. 1978. *Scientific progress. A philosophical essay on the economics of research in natural science,* Oxford: Blackwell.

Richtmyer, F.K. 1932. 'The romance of the next decimal place', *Science, 75,* 1-5.

Schilpp, P.A. 1974. (Ed.), *The philosophy of Karl Popper,* 2 vols., La Salle, Illinois: Open Court.

Shewhart, 1931. *Economic quality control of manufactured product,* London: MacMillan.

Wolenski, J. and Köhler, E. 1999. (Eds.) *Alfred Tarski and the Vienna Circle,* Dordrecht: Kluwer.

Wolf, C. 1889-1891. (Ed.) *Mémoires sur la pendule, précédés d'une bibliographie,* 2 vols., Paris: Gauthiers-Villars.
Worrall, J. 1989. 'Why both Popper and Watkins fail to solve the problem of induction', in F. d'Agostino and I.C. Jarvie (eds.), *Freedom and rationality,* Dordrecht: Kluwer, 257-296.
Ziman, J. 1978. *Reliable knowledge. An exploration of ground for belief in science*, Cambridge: Cambridge University Press.

Further references

The works listed here have not been cited above, but they valuably guided the considerations presented.

Fleck, L. 1935. *Entstehung und Entwicklung einer wissenschaftliche Tatsache: Einführung in die Lehre vom Denkstil und Denkkollectiv,* Basel: Schwabe. [English trans.: *Genesis and development of a scientific fact*, Chicago: University of Chicago Press, 1979.]
Gjertsen, D. 1989. *Science and philosophy. Past and present*, London: Penguin.
Hacking, I. 1983. *Representing and intervening. Introductory topics in the philosophy of science*, Cambridge: Cambridge University Press.
Kitcher, P. 1993. *The advancement of science. Science without legend, objectivity without illusions*, New York and Oxford: Oxford University Press.
Maxwell, N. 1998. *The comprehensibility of the universe,* Oxford: Oxford University Press.
Medawar, P.B. 1969. *Induction and intuition in scientific thought*, London: Methuen.
Perutz, M. 1991. *Is science necessary?*. New York and Oxford: Oxford University Press.
Richards, S. 1987. *Philosophy and sociology of science. An introduction*, 2nd ed. Oxford: Blackwell.
Waddington, C.H. 1977. *Tools for thought*, London: Cape. [Paperback reprints.]

Middlesex University at Enfield
Middlesex EN3 4SF
England
Letters to:
43, St. Leonard's Road, Bengeo, Herts.
SG14 3JW, U.K.
ivor2@mdx.ac.uk

LADISLAV KVASZ

HOW CAN A FALSIFIED THEORY REMAIN CORROBORATED?

Comments on I. Grattan-Guinness's "The Place of the Notion of Corroboration in Karl Popper's Philosophy of Science"

Coming from a mathematical background, I was always puzzled by Popper's view, according to which, after the falsification of a scientific theory its degree of corroboration becomes zero. Most of the scientific theories taught in the physics departments have already been falsified, and what is the point of teaching theories, whose degree of corroboration is zero? The first important observation to make is that not all cases of falsification are the same. In some cases, as for instance in the case of the theory of phlogiston, the falsification happened in agreement with the Popperian picture. Scientists discarded the falsified theory and opted for its alternative. Therefore, nowadays nobody tries to make a scientific contribution to the theory of phlogiston. Nevertheless, there are cases, and Newtonian mechanics is surely the most important among them, when the behaviour of the scientists after the falsification of the theory is from the Popperian point of view incomprehensible. Many scientists were not ready to discard the falsified theory, and not for irrational reasons. Some of the deepest discoveries in Newtonian mechanics – as for instance the famous Kolmogorov, Arnold, Moser theorem – were made many years after this theory was falsified. I think that Andrej Kolmogorov, Vladimir Arnold or Jürgen Moser cannot be compared to an Aristotelian philosopher, who adheres to his pet theory after its falsification. What these three mathematicians did was a fundamental contribution to modern science. And they did it despite the fact that the theory, to which they made their contribution, was already falsified. Thus, it seems to be necessary to introduce some distinctions into the Popperian theory of falsification, in order to capture the differences between the falsification of the theory of phlogiston and the falsification of Newtonian mechanics. The situation here is similar to Kuhn's theory of scientific revolutions (see the introduction to Gillies 1992, or Kvasz 1999).

But let us put the question of possible refinements of the Popperian theory of falsification aside – our problem here is the problem of corroboration. I will stick to the case of the Newtonian mechanics, the case discussed by Professor Grattan-Guinness. What are scientists like Kolmogorov doing when they work on a theory whose degree of corroboration is zero? The first answer might be that actually they are not developing Newtonian mechanics (N), which has been falsified, but rather a new theory (N'), obtained from (N) by the introduction of some re-

F. Stadler (ed.), Induction and Deduction in the Sciences, 263–271.
© 2004 *Kluwer Academic Publishers.*

strictions, as $v << c$. This new theory (N') may remain still highly corroborated, because actually all cases of corroboration in mechanics happened in situations where the velocities of the bodies were relatively small. Therefore developing further the theory (N') might be a rational choice.

Unfortunately, the situation is not so simple. We cannot add the condition $v << c$ to the principles of Newtonian mechanics, because it would lead to contradictions. Newtonian mechanics contains rules for the addition of velocities. In some situations, these rules allow us to derive velocities much higher than the speed of light and these derivations would come into contradiction with the new rule $v << c$. Therefore, this restriction cannot be introduced on the object level of Newtonian mechanics. Another possibility is to try to introduce the rule $v << c$ on the meta-level, as a methodological rule, saying that whenever we derive a proposition containing velocities of the same magnitude as the speed of light, we have to drop this proposition. But it seems that this does not work either. It would break down the closeness of the theory with regard to the relation of logical consequence, which is a substantial part of the definition of the notion of a scientific theory. By adding the rule $v << c$ to the meta-level Newtonian mechanics would turn into a fragment, a set of propositions which is not closed under the relation of logical consequence, and therefore it would not be a scientific theory. Thus, we see that neither of the two ways of creating the theory (N') is feasible.

If we open some contemporary exposition of Newtonian mechanics, as Arnold (1974) or Abraham and Marsden (1978), we find that no reinterpretation is actually adopted. These books develop simply the theory (N), the classical Newtonian mechanics, without any attempt to introduce any changes into it in order to come to terms with the fact of its falsification. And it is good so, because the only way how to adopt the structure of classical mechanics to the relativistic phenomena, which led to its falsification, is to do what Einstein did. That means to change in a radical way all the basic notions of the theory. The only consistent answer to the falsification of Newtonian mechanics is the relativistic mechanics. Nothing in-between is logically consistent. The notions and propositions of scientific theories (at least in theoretical physics) are too interwoven, and, therefore, any local patchwork simply does not work. There are only two consistent options: the Newtonian or the Einsteinian.

The restricting rule $v << c$ can be seen rather as a correspondence principle, a meta-theoretical rule of the relativistic mechanics, discriminating a *fragment of the relativistic theory*, which in many respects is similar (or corresponding) to Newtonian mechanics. The existence of this fragment explains the success of Newtonian mechanics. Newtonian mechanics was successful in those cases, where it was restricted to small velocities, because in that area the Newtonian calculations gave practically the same results as (or more precisely were the limiting cases of) the relativistic calculations. It is important to realize that the rule $v << c$ determines only a fragment, and not a theory, and that this fragment is a part of the relativistic and not of the Newtonian mechanics. A widespread mis-

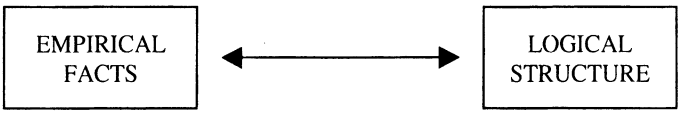
take consists in transferring the rule $v << c$ from the relativistic mechanics (to which it belongs, because the speed of light is a relativistic constant) to the Newtonian mechanics, and to interpret it as a condition under which Newtonian mechanics is still valid. There is no doubt that Newtonian mechanics is valid for small velocities, but the proposition saying this is a proposition *about* Newtonian mechanics and not a proposition *of* Newtonian mechanics. It is a proposition of the relativistic mechanics, because when we refer to the boundaries of Newtonian mechanics, we are already standing outside of it.

Thus, to sum up, we have seen that after the falsification of Newtonian mechanics in 1919, this theory was neither abandoned, nor replaced by some theory (N'). Scientists like Kolmogorov, Arnold, and Moser were developing the old, falsified Newtonian mechanics. This seems to contradict Popper's intuition that the degree of corroboration of a falsified theory is zero. So his definition of the notion of corroboration needs some revision. Our suggestion is that after its falsification Newtonian mechanics was not discarded, but rather transformed into a mathematical theory. This view is substantiated by the fact that Kolmogorov and Arnold are mathematicians. Thus, a scientific theory, after it gets falsified, can be either abandoned (and replaced by another theory as in the case of the theory of phlogiston), or transformed into a mathematical theory. We believe that the latter is what happened with Newtonian mechanics. Therefore, we have to look more closely on the relation between mathematics and physics in the history of scientific thought.

1. A SCIENTIFIC THEORY AND ITS LINGUISTIC FRAMEWORK

Classical positivism had a strongly anti-Kantian attitude. While Kant overestimated the role of mathematics in science, positivists like Mach underestimated it. In logical positivism this general tendency of underestimating the role of mathematics in science was strengthened due to the adoption of logicism as a philosophical interpretation of mathematics. Logicism is one of the alternative approaches to the foundations of mathematics, but it is definitely not the one adopted by working mathematicians. The majority of working mathematicians are Platonists, many of them are formalists, some of them are intuitionists, but I have not met a mathematician who would be a logicist. I think this is so because logicism is lacking understanding of mathematical practice. Logicism is a view of the foundations of mathematics, which was developed mostly by philosophers and logicians. But Newton, Euler, Lagrange, Hamilton, Liouville, Noether, Poincaré, or Kolmogorov, to name only the most famous mathematicians who made fundamental contributions to Newtonian mechanics, definitely were not logicists. Therefore, to try to understand their contributions within the frame of logicism creates more problems than it can solve. Many difficulties in the philosophy of science are at least partially due to the oversimplified picture of science in which mathematics is reduced to logic. For instance, Carnap's discussion of inductive

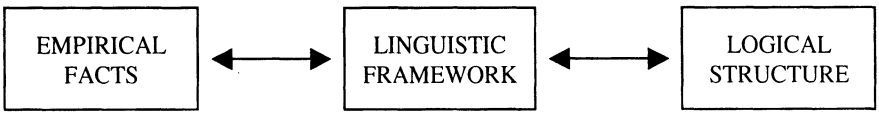
verification of scientific theories can be seen as an implicit denial of the role of mathematics in science. According to logical positivists, mathematics is analytical, and therefore it contributes nothing relevant to the content of a scientific theory. So the content of a scientific theory has to be captured solely in terms of empirical procedures and logical analysis, i.e., without any recourse to mathematics. In his criticism of Carnap's theory of induction, Popper took over Carnap's general framework, and so even though he turned from verification to falsification, and destroyed (or believed to have destroyed) the myth of induction, he shared Carnap's neglect of mathematics. Thus, Popper's theory of corroboration is formulated within the following scheme:



Let us now go back to the paper of Professor Grattan-Guinness. In discussing Popper's concept of corroboration, Professor Grattan-Guinness introduced two topics which could shed light on the question of the role of mathematics in coordinating empirical facts with the logical structure of a scientific theory. One of these topics concerns technology, the other one education. If we consider the ways in which Newtonian mechanics *is used* today by technicians or *is taught* at schools of different levels, we can see that in Newtonian mechanics there is a series of transitory layers between the empirical facts and the logical structure. What technicians use and what teachers teach are not just empirical facts and their logical reconstruction. Instead, we find here a wide range of different mathematical proofs, derivations and mere calculations taking place on different levels of abstraction, that reach from elementary calculus to the theory of manifolds. In order to capture these layers, I suggest that we discriminate between a *scientific theory* and its *linguistic framework*. If we look on Newtonian mechanics, as it is used by technicians, we discover that they are not interested in planetary motions, the trajectories of the comets or the theory of tides, which were the areas of principal corroboration of the theory in Newton's times. Similarly, they might not even know about the motion of the perihelion of the Mercury, which was where Newton's theory was falsified. They are interested in techniques of solving specific problems that occur in their particular field. Thus, it is fair to say, that the technicians are interested in the linguistic framework of Newtonian mechanics. It is this linguistic framework that the physicists teach their students while lecturing on Newtonian mechanics. Technical applications and science education focus on Newtonian mechanics not as a scientific theory, i.e., as a set of true propositions describing the world, but rather as a linguistic framework, i.e., a suitable tool which can be used in many particular situations.

On the background of this discrimination, we can reach a better understanding of Newtonian mechanics as it is developed today in mathematical physics. As we have already stated, instead of abandoning Newtonian mechanics after its falsification, scientists transformed it into a mathematical theory. Now we can give a precise formulation of this view. A theory is a system of propositions closed under the relation of logical consequence. The main difference between Newtonian mechanics understood as a scientific theory and Newtonian mechanics taken as a mathematical theory is that in mathematics we do not confront the propositions with empirical observations, but rest them on axioms. Nevertheless, by turning Newtonian mechanics into a mathematical theory its relevance for science is not totally lost, because *mathematics is the language of scientific theories*. Thus Newtonian mechanics, taken as a mathematical theory, can serve as a linguistic framework for many branches of science, for instance in the description of the measurement devices, in engineering or in architecture. To stick to the example of the car crash, introduced by Professor Grattan-Guinness, I can use my car (or the linguistic framework of a theory), although there are situations in which I know that a crash can occur. I will simply steer my car (or my calculations) away from that dangerous spot. Thus, although I know that there is a situation, where the theory gets falsified, it still can be used as a linguistic framework in all those situations where it does not.

The notion of linguistic framework is relevant not only in order to understand why it is rational to develop, to teach and to rely on a falsified theory. It can help us to understand the process of discovery of new theories as well. In order to develop his *Principia*, Newton had to introduce a new branch of mathematics, the calculus. Thus, the transition from the empirical facts to a scientific theory, which explains them, was not direct. It was mediated by the linguistic framework. Unfortunately, Newton suppressed a great deal of this framework, and he formulated his theory using synthetic geometry. Nevertheless, the frequent use of limit transitions shows that his theory was based on his new mathematics. In our view, therefore, a more adequate scheme for the articulation of many problems in the philosophy of science would be the following one:



This scheme opens a new view on the problem of induction. I do not want to raise the question of induction in general, but rather to restrict myself to Newton, who claimed that he had *proved the law of universal gravitation by induction from the phenomena*. What did he mean by this? If we try to understand Newton's words on the background of the first scheme, we end up in serious difficulties. It is clear that Newton did not use any inductive logic or probabilistic

arguments. Therefore, we cannot reconstruct Newton's discovery using inductive logic (whatever it might be) or Bayesian theory, for the simple reason that in the *Principia* there are no traces of such devices. Nevertheless, if we interpret Newton's words from the point of view of our second scheme, we can give them rational meaning. I believe that Newton's proof by induction from the phenomena can be understood as a reasoning that is closer to mathematical induction than to inductive logic.

In the case of mathematical induction coming from

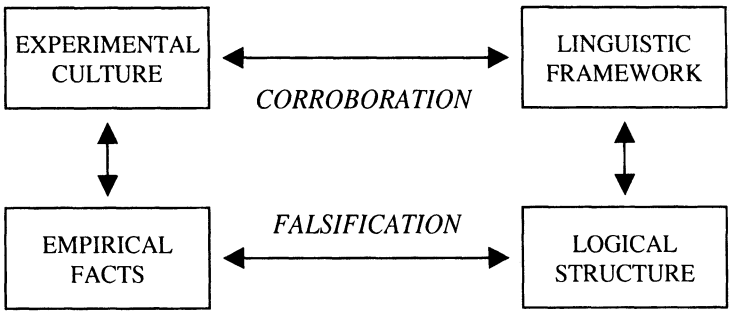
$$f(1) \text{ and } (\forall n)[f(n) \Rightarrow f(n+1)]$$

to the conclusion that $(\forall n) f(n)$, we use not only deductive logic, but we also make use of the well ordering of the natural numbers. I believe that the basis of Newton's proof was the *analytic structure* of the calculus, similarly to the *ordering structure* of natural numbers being the basis of mathematical induction. Popper was totally correct in that relying purely on logic we cannot infer any general law from finite empirical data. There are an infinite number of logically possible curves passing through these data. If we view this problem not within the scheme used by logical positivists, dividing science into empirical facts and logical structure, but within our second scheme, taking into account the mediating role of mathematics, the picture changes fundamentally. Instead of all logically possible functions, we can consider all analytically suitable ones. Thus, analogically to the well ordering of natural numbers, there is an ordering of the system of analytic functions (for instance with respect to their smoothness or their degree). In this way, calculus introduces a structure into the data, and it is this structure on which Newton's inductive proof can rely. One nice example of such reduction of possibilities is discussed by Arnold. It can be proved that only in two cases of central forces, $F(r) \approx \frac{1}{r^2}$ and $F(r) \approx -r$, bodies move on closed trajectories. Therefore, from the empirical fact that the trajectories of the planets are closed, we can infer that the gravitational force is either proportional to the distance or inversely proportional to its square. Nevertheless, similarly to the mathematical induction, the inference here is a deductive one. Analogically, Newton's proof by induction was not based on some new kind of logic. It was an ordinary deductive argument, but it was imbedded in a mathematical structure. This simple example shows that mathematics plays a fundamental role in science, which role cannot be reduced to logic. Mathematics forms the linguistic framework, which offers structures (of ordering, symmetry, topology, differentiation, measure), by help of which it is possible to relate the data to the theoretical propositions. For a more detailed discussion of the linguistic aspects of mathematics see Kvasz (2000). Thus, our first response to the problems of the relation between corroboration and falsification could be stated as follows: What was falsified by the discovery of the anomaly in the rotation of the perihelion of Mercury was Newtonian mechanics as a scientific theory, i.e., as the true picture

of our physical world. On the other hand, what we corroborate in our everyday life is Newtonian mechanics as a linguistic framework, i.e., as a tool for organizing our everyday experience.

2. AN ALTERNATIVE INTERPRETATION OF THE NOTIONS OF FALSIFICATION AND CORROBORATION

A fundamental shift in philosophy – the *linguistic turn* – occurred a few years after the famous Carnap-Popper controversy. Unfortunately, the *linguistic turn* turned out to be a turn towards ordinary language. Thus, similarly to Carnap and Popper, who ignored the mathematical language used in the development of scientific theories, and reduced the whole discussion to logic, the *linguistic turn* also ignores the mathematical language. The only difference is that now mathematical language is not replaced by logic, but by ordinary language. Thus. what is still missing is a turn to the language of science itself. So let us take this turn. In a similar way, as the propositions of a theory are embedded in its *linguistic framework*, the empirical facts are embedded in something what we suggest to call *experimental culture*. A great deal of this experimental culture is reconstructed in contemporary history of science. Historians are engaged in reproducing the particular experiments of a given historical period, using, as far as possible, the original equipment, materials and procedures, in order to see what the scientists of the past did observe. Thus, our final scheme for the reconstruction of the development of science looks like this:



In this scheme, the *falsification* takes place on the level of empirical facts and logical structure, while the *corroboration* works on the level of experimental culture and linguistic framework. So we can characterize corroboration as a *stable equilibrium between the experimental culture and the linguistic framework of a scientific theory*. If we focus on the historical details, we find that most of the falsifications in science occurred as a result of changes in the experimental culture or in the linguistic framework. For example, the famous Michelson-

Morley experiment made use of very sensitive interferometers, i.e., equipment that did not belong to the experimental culture of classical mechanics. If we make the experimental culture and the linguistic framework parts of our reconstruction of science, we will be able to analyse the innovations in these areas. An innovation in the *experimental culture* (e.g., the invention of the interferometer) might produce new, hitherto unimaginable empirical facts, which might lead to the falsification of the theory. Similarly, an innovation in the *linguistic framework* (e.g., the introduction of Riemannian differential geometry) might lead to a refinement of the theoretical predictions and comparing them with the already existing data might eventually result in a falsification of the theory itself. Thus, changes in the experimental culture or in the linguistic framework may disturb the equilibrium between the theory and the realm of empirical facts.

Nevertheless, if we consider the linguistic framework of classical mechanics (say, in the form developed by Lagrange and Hamilton) and relate it to the experimental culture that belongs to this theory (and which is still used in technical applications), a stable equilibrium can be established. We call this equilibrium *corroboration*. It is a result of centuries of testing, and it cannot be destroyed by a simple falsification of the theory. The actual falsification of classical mechanics took place outside its experimental culture. After some time, the new experiments, which were able to display relativistic phenomena, created a new experimental culture, which reached equilibrium with the theory of relativity. Thus, not only the falsified theory was replaced by a new theory, but also its experimental culture was replaced by a new one. Nevertheless, it is important to realize that after the falsification of a theory, its linguistic framework and experimental culture survive. The experimental culture survives in technical applications and its linguistic framework survives in education. Therefore, the corroboration, interpreted as a stable equilibrium between these two aspects of the theory, remains mostly unaffected by the actual falsification. We believe that scientists are not so bold in proposing new theories and not so severe in testing them, as Popper believed. The progress of science is a much more prosaic enterprise. It consists rather in a systematic development of the linguistic framework and in a systematic refinement of the experimental culture, and not in mere bold guesses and severe testing. Of course, if we leave the linguistic framework and the experimental culture outside the picture, what remains is the Popperian picture of boldness and severity.

REFERENCES

Ralph Abraham/Jerrold Marsden, *Foundations of Mechanics*. Reading, Mass.: Benjamin/Cummings 1978.

Vladimir I. Arnold, *Mathematical methods of classical mechanics*. New York: Springer 1974.

Donald Gillies (Ed.), *Revolutions in Mathematics*. Oxford: Clarendon Press 1992.

Ladislav Kvasz, „On Classification of scientific revolutions", in: *Journal for General Philosophy of Science* 30, 2, 1999, pp. 201-232.
Ladislav Kvasz, „Changes of Language in the Development of mathematics", in: *Philosophia mathematica*, 8, 1, 2000, pp. 47-83.

Department of Humanities
Faculty of Mathematics, Physics and Informatics
Comenius University
Mlynska dolina
84248 Bratislava
Slovak Republic
kvasz@fmph.uniba.sk

KARL MILFORD

INDUCTIVISM IN 19TH CENTURY GERMAN ECONOMICS

1. ANTI- AND PRO-NATURALISTIC POSITIONS

In his *The Poverty of Historicism*[1] K.R. Popper and before him F. Kaufmann[2] distinguish two broad classes of epistemological and methodological positions held in the social sciences: Anti-naturalistic positions and pro-naturalistic positions. These positions are distinguished with respect to their attitude regarding the applicability of the methods of the natural sciences, or rather what the representatives of the anti- and pro-naturalistic positions assume to be the method of the natural sciences. According to Popper and Kaufmann the representatives of anti-naturalistic positions hold that the methods of the natural sciences can not be applied in the social sciences, whereas the representatives of the pro-naturalistic positions have the opposite view. This of course raises the question what views the representatives of the anti- and pro-naturalistic positions embrace with respect to the methods of the natural sciences. It is interesting to note, however, that Popper in his *The Poverty of Historicism*, indicates those views only briefly, the reasons presumably being the special tasks and aims he set himself in that analysis. These are primarily to show the disastrous political and social consequences of a false social science methodology and the invalidity of the ideas of historical necessity and that of laws of historical development. But it is well known that Popper analysed and discussed the methods and the philosophy of science with great rigour and acumen in his other books and articles and that he triggered a philosophical revolution in that field.[3] Before Popper the prevailing view with respect to the methodology of the empirical sciences was that the natural sciences are characterised by the so-called method of induction and that the laws of nature are inductively inferred by content-enlarging and truth-preserving inferences from an empirical basis that is absolutely certain because its truth is proven.[4] In that pre-Popperian view empirical science is demarcated by the method of induction from the results of other areas of human intellectual activity such as logic, art or philosophy that do not claim to propose laws of nature that presume to be strictly universally valid and empirical. Yet it was well known that the so-called method of induction was logically invalid and triggered all sorts of different and pressing epistemological and methodological problems that called for urgent resolution.[5] Since it seemed that unless those problems were solved the rationality of science could no longer be defended many authors tried to find

F. Stadler (ed.), Induction and Deduction in the Sciences, 273–291.
© 2004 *Kluwer Academic Publishers.*

acceptable solutions for the two fundamental problems of epistemology, the problem of induction and of demarcation.[6] However, Popper has shown that apart from his, all the suggested solutions are invalid and that some of them even implied self-defeating and irrational consequences, which their authors desperately tried to avoid when suggesting that solution. [7]

With respect to the suggested solutions of the problem of induction (and thereby that of demarcation) Popper distinguishes two broad classes: inductivist and deductivist solutions.[8] He points out that inductivist solutions share the general premise that science starts with single and unprejudiced observations and that these observations including perhaps introspection as well constitute the so-called empirical basis. Many of those inductivist positions, however, also share the view, that in order to apply the method of induction successfully it is necessary to have the possibility of making repeated observations. Without that possibility they believe that no regularity could be found and verified. In their view the possibility of making such repeated observations shows itself most prominently in the natural sciences and in particular in the possibility of making controlled experiments because the natural sciences study a "universe" that is characterised by regularity or uniformity. This logical feature of the natural universe guarantees the success of the natural sciences and is regarded as a precondition for applying induction successfully.[9]

It seems that the representatives of the anti- and the pro-naturalistic positions alike share those inductivist views about the methodology of the natural sciences. However, they disagree with respect to the question of the possibility of making repeated observations for the social sciences. Whereas they agree that this possibility exists to a very high degree for the natural sciences, because the "natural universe" is ontologically characterised by the so-called "uniformity of nature", they emphasise that the "social universe" is much more liable to change. From this it is concluded that the method of induction cannot be applied in the social sciences as rigorously as in the natural sciences. However, opinions diverge with respect to what degree this possibility exists.[10] Anti-naturalistic positions usually emphasise "change" as the prime ontological characteristic of the "social universe" and tend to be rather sceptical with respect to the possibility of making repeated observations in the social sciences. But although they tend to deny the possibility of making experiments and emphasise the historicity, the individuality and the newness of social phenomena, they do not exclude the applicability of inductivist methods in the social sciences altogether. Some of the anti-naturalistic positions hold that the empirical basis of the social sciences is constituted by the history of mankind and that from this basis "laws of lesser strictness" than those of the natural sciences can be induced. In their view laws of that kind claim to be valid for certain historical periods only, such as the laws of feudalism or of capitalism. Other anti-naturalistic positions emphasise the complexity of social phenomena or processes and that of the social universe. In their view this complexity is the result of the interplay of a vast amount of different causes and causal laws and that this vast amount of the interplay of physical, biological

and psychological laws prevent any possibility of making repeated observations directly. Although this seems to render the application of inductive methods and thus the attainment of genuine scientific social knowledge impossible, some representatives of those anti-naturalistic positions believe that this difficulty can be overcome. In their view it is possible to reconstruct social processes and facts from laws of other disciplines which have already been proven inductively, as for instance the laws of empirical psychology, since those disciplines scrutinise a "universe" that is characterised by repeatability and thus renders the method of induction applicable.[11] With respect to "genuine newness" or "spontaneity" it is held that similar methods could be applied, i.e. reconstructing genuine new social facts as an outcome of inductively proven so-called heteropathic and psychological laws.[12] Some other anti-naturalistic positions try to circumvent the seeming impossibility of applying inductive methods in the social sciences by pointing out that the task of the social sciences is to uncover the essence of social institutions and processes. The Aristotelian version of that view holds that the essences of things lie as potentiality within the things themselves and come to existence in the actual historical process. Observable social institutions are emanations of essences or spirits. In order to achieve genuine social knowledge the task of the social scientist is to observe the historical development of institutions or peoples and by discovering repeated developments to uncover the laws of historical development which describe the essence of those institutions.[13]

In contrast to the anti-naturalistic positions, pro-naturalistic positions tend to hold that the social universe is characterised by a higher degree of repeatability and that therefore the methods of induction are more directly applicable. It is assumed that laws of historical development can be found through direct observations of the development of different peoples or institutions, or that structural laws can be found according to which one historical period is replaced by another one. The possibility of experiments is not denied but pro-naturalistic positions do not regard the possibility of artificially repeated observations to be very important in the social sciences. They hold that similarly to the natural sciences the social sciences need a proven and absolutely certain factual basis upon which the social theories are built. This factual or empirical basis is the history of mankind and from this basis social theories are to be inferred by inductive methods. According to this view price theory is to be based upon detailed historical studies that provide the data from which that theory is inferred; business theories alike are to be based upon a vast amount of data and so forth. Other versions of pro-naturalistic positions regard Newton's celestial mechanics as a model of the social universe and aim to uncover social regularities or laws between the social sectors of an economy by scrutinising data.

Although the social sciences' epistemological and methodological literature of the 19th and beginning 20th century discussed the method of induction primarily in terms of its possible application in the social sciences and took the validity of that method somehow for granted, many authors were aware of the logical and methodological problems of that method. They believed that the logical and

epistemological problems of induction which were triggered by Hume's logical objection against content-enlarging and truth-preserving inferences needed resolution before discussing its possible application in the social sciences. They discussed the problem of induction and tried to find a logically and epistemologically valid solution for the problem of establishing empirical and strictly universal statements by content-enlarging and truth-preserving inferences. Some authors tried to find some kind of induction principle and sought to establish that principle on a priori grounds or by experience. Other authors sought to solve that problem by introducing additional truth values in the form of probabilities that could be assigned to statements. And other authors held that social and economic theories are systems of implicit definitions that were useful or not in order to make forecasts, but neither true nor false. However, the social sciences' epistemological and methodological literature of the 19th and beginning 20th century discusses not only the problems of a possible application of inductive methods in the social sciences and their logical validity, but other problems that were closely related to them. Another major problem discussed in that literature is that of the structure of a satisfactory explanation in the theoretical social sciences. In general one can distinguish two positions here: individualistic and collectivistic positions. Individualistic positions hold the view that satisfactory explanations in the theoretical social sciences explain social institutions, processes or events as the unintended outcome of the interplay of the individuals' intended actions. According to that view the intended actions of humans are to be explained by modelling social situations that describe some of the major social conditions under which humans act as well as their aims, preferences and knowledge. By putting together models that explain the plans and the decisions of individuals it was thought that social institutions or processes such as the working mechanism of different markets could be explained.[14] In that context some authors emphasised that these social science explanations or models do not involve any psychological considerations[15] while others thought that social science knowledge ought to be derived from psychological laws, that are inductively inferred from internal (introspection) and from external observation.[16] Since the individuals' actions are to be explained in terms of their preferences, aims and knowledge, the positions of (methodological) individualism may also be described as the view that satisfactory explanations of the theoretical social sciences explain social institutions or processes as the unintended result of the interplay of individuals' evaluations of social situations. In contrast to that position, the representatives of (methodological) collectivist, or as one might also call them, inductivist essentialist positions hold that the methodological requirements of (methodological) individualism cannot be fulfilled even in principle. The representatives of that position emphasise that individuals act within a social and cultural framework or tradition and that satisfactory explanations in the social sciences first need to provide a satisfactory account of that framework regarding its historicity and development before explaining individual actions and their results. However, it is emphasised that an account of that framework cannot be

given in individualistic terms, because individuals act within that framework and an individualistic explanation of it would lead to an infinite regress. If the social and cultural framework in which humans perform were to be explained as an unintended outcome of the individuals' evaluations, one would have to assume a second social and cultural framework in order to explain the first one and so ad infinitum. As a possible solution of that problem the representatives of methodological collectivism or methodological inductivist essentialism suggest uncovering the laws of historical development of that social and cultural framework or proposing laws that describe regularities between different parts of that framework, by considering history as an empirical basis from which they are inferred.

For the sake of contrasting these two methodological positions one may summarise their basic methodological principles. Proposing satisfactory explanations in the social sciences the positions of methodological individualism require to "propose theories or social laws by reconstructing them as the unintended outcome of the interplay of the individuals' evaluations of social situations". In contrast the positions of methodological inductivist essentialism call for satisfactory social science explanations to "propose laws of historical development of the social and cultural framework or laws that describe regularities between parts of that framework in which individuals act". Both positions assert something about the possibility and the kind of structure a satisfactory explanation in the theoretical sciences ought to have. According to methodological individualism it is in principle always possible to explain social institutions, processes or events as the unintended result of the interplay of the individuals' evaluations of social situations and that false explanations are due to the scientist's imperfect knowledge and not to fundamental scientific and methodological reasons. In contrast inductivist methodological essentialism claims that there is at least one social fact that for principled scientific reasons cannot be explained as the unintended result of the individuals' evaluations of social situations because an attempt to explain the preconditions of individual evaluations on an individualistic basis leads to an infinite regress.

2. METHODOLOGICAL AND METAPHYSICAL POSITIONS

Since the positions of methodological individualism and methodological inductivist essentialism assert something about science these positions could also be called scientific individualism and scientific inductivist essentialism. These positions, however, have to be distinguished from their metaphysical counterparts which assert something different, namely something about the ontological structure of the "universe" under investigation. They do not assert something about science but something about the functioning of the world and its structure and thus propose some kind of ontology. The positions of metaphysical essentialism suggest that the social universe is inhabited by holistic entities that

possess a nature or an essence as potentiality, which – in the Aristotelian version of that position comes into existence in the course of their historical development. According to this view there exist social and cultural institutions or entities that are in fact not the result of the interplay of the individuals' evaluations of social situations or of their intended actions. In contrast, metaphysical individualism holds that such holistic entities do not exist and that only the individuals' actions and their structural outcomes are real. In this view social institutions are in fact the unintended consequences of individual intended actions.[17]

Although the metaphysical and methodological positions of individualism and inductivist essentialism make different propositions, one asserting some kind of ontology, the other one asserting something about science, a relation exists between them, since methodological positions may be regarded as reinterpretations of their metaphysical counterparts. In the epistemological and methodological literature of the social sciences this relation is important not only for the critical assessment of different positions because insoluble metaphysical problems become soluble if their methodological reinterpretation is considered. It is of importance also because some authors try to defend and support their methodological positions by referring to their metaphysical counterparts. As mentioned above some representatives of anti-naturalistic positions defend their rejection of a direct application of inductive methods in the social sciences by emphasising that, in contrast to the social universe, the natural universe is characterised by a regularity or conformity that can be ascertained through repeated observations. The possibility of making repeated observations is regarded by them as a precondition for the application of inductive methods, a precondition that, however, does not apply to the social universe and makes the direct application of the inductive method either impossible or possible only to a limited extent. Here a synthetic and unfalsifiable metaphysical assertion is used either to support or to reject a methodological position. Arguments of that kind are of some importance for instance in the works of J.S. Mill, A. Marshall and W. Roscher, but Menger already showed their futility. The metaphysical assertion "There exists regularity" cannot solve the methodological problem of justifying an induction principle.[18] Yet these attempts can be explained if methodological positions are regarded as reinterpretations of metaphysical ones. Such attempts for instance sought to derive the task of social science, i.e. a methodological proposition, from a metaphysical one about the structure of the "universe". The methodological proposition of methodological inductivist essentialism that the task of the theoretical social sciences is to propose the laws of historical development is justified by the metaphysical proposition of the existence of holistic entities and their essences. Since the representatives of metaphysical essentialism believe in the existence of holistic entities and their essences that come into being in the actual historical process they declare that it is the task of the theoretical social sciences to uncover the laws of historical development because they describe the essence of social institutions and therefore deliver genuine scientific social knowledge. Similarly the representatives of individualistic positions tend

to justify the methodological proposition that it is the task of the theoretical social sciences to posit social laws in the form of theories or models that generate those regularities as the unintended consequences of the interplay of the individuals' intended actions by asserting that the social universe is indeed the unintended result of the individuals' evaluations of social situations.

However, the metaphysical positions cannot really support the methodological positions here.[19] For, unless logical objections to such a procedure exist, the search for regularities between social aggregates may be quite reasonable; and to explain social institutions and processes as the unintended outcome of the interplay of the individuals' evaluations of social situations, even in a social universe whose regularities are not the unintended outcome of the individuals' intended action, may be quite reasonable as well, if no other method exists that provides better explanations of social facts. Both metaphysical positions assert the existence of regularities as a precondition for uncovering laws, but the search for social laws or regularities does not presuppose such a precondition.[20]

3. METHODOLOGICAL DISCUSSIONS AND THE DEVELOPMENT OF ECONOMIC THEORY

Although some of the methodological and epistemological problems discussed in the social sciences are fairly autonomous, such as logical ones, it should be noted that many of them are closely related to the development of social theory itself. The history of a discipline is primarily one of problems and problem situations and often some open explanatory problem of a certain discipline leads to the discussion of methodological and epistemological issues. It is then hoped that by solving the latter ones, solutions for the former will be found. This is particularly manifest with respect to the development of price theory in 19th century German economics. In general, many 19th century economists, in particular German authors, criticised the in their view considerably low explanatory power of economic theories and thought them to be too abstract and unrealistic in order to deliver satisfactory explanations of social facts and a good understanding of the working mechanism of social and economic systems.[21] But the most urgent problem that according to the German authors needed resolution was the explanation of the determination of relative prices and of exchange ratios. It was assumed that the prevailing labour and cost theoretical explanations of relative prices and exchange ratios were fairly unsatisfactory because they could not provide a proper explanation of exchange and yielded a heterogeneous and inconsistent price theory. Price theory, however, was regarded as the corner stone of every economic theory and therefore the state of the art was considered to be most unsatisfactory. The resolution of the open problems of price theory were therefore regarded as most important in order to raise the explanatory power of economic theories.

The critics of labour and cost theoretical approaches believed that this theory was a major source of the low explanatory power of economic theory and its unrealisticness. It was assumed that an answer to all those open problems could be found if economic theory were to stick to and regard facts more faithfully. In their view the premises of (the Ricardian) models did not describe social situations realistically, so that it seemed quite obvious that the explanations or forecasts derived deductively from such models had to be poor or even completely false by necessity, even if derived correctly.[22] However since deductive reasoning is purely analytical, the question of an improved empirical content of economic theories seemed to be possible only if the truth of the principles and premises used in economic models were verified by experience. The question of improving the explanatory power of economic theories thus seemed to be one of method, and many authors eventually thought that that method was induction. For in contrast to the analytical character of deductive inferences the content-enlarging and truth-preserving inferences of the inductive method seemed to promise growth of genuine scientific knowledge.

Basically the problem situation that prevailed in 19th century economics may be described by four major elements. On the one hand by two behavioural theories, the theory of objective evaluations and the theory of subjective evaluations that were designed to explain relative prices or exchange ratios and on the other by two methodological theories, methodological individualism and methodological inductivist essentialism or collectivism and their metaphysical counterparts, respectively, that were designed to define the structure of satisfactory explanations in economics. Since the methodological positions and their metaphysical counterparts have already been described only an account of the two behavioural theories has to be given to provide a complete description of the situation.

Whereas the methodological positions assert something about science and their metaphysical counterparts propose some kind of ontology the two behavioural theories provide different explanations of how individuals evaluate commodities or social situations in order to explain exchange ratios or relative prices. According to the objective theory of objective evaluations individuals evaluate commodities or social situations in general according to some objectively given common standard, such as labour inputs measured in hours. However, there exist different versions of that theory. Labour cost theories explain the exchange ratios or relative prices of commodities by the ratio of labour inputs measured in hours that are needed to produce them. According to that view commodities exchange according to the ratio of their labour inputs. Other versions of that theory, however, explain the exchange ratios and relative prices of commodities by the ratio of how much labour these products can command on the market. And although the representatives of either versions of such labour theories realised that they suffer from severe difficulties if capital is introduced as a factor of production they thought like Ricardo that labour theories provided a good approximate explanation of the ratios according to which commodities exchange on markets.[23] In contrast to the theories of objective evaluations the theories of subjective

evaluations hold that individuals evaluate commodities or social situations according to their subjective preferences or utility considerations only. In this view the ratios according to which commodities exchange, or the relative prices, are determined by the marginal utilities – that is, by the increase in utility an individual gains from an additional unit of a commodity given the quantity or stock of that commodity already owned by that individual. The kind of physical things or action individuals regard as commodities or services are determined exclusively by subjectivist preferences even if the technical characteristics of those things or actions are objectively unsuitable for achieving the aims the individuals want to achieve with them. According to that view the ratios according to which commodities and services exchange on markets are determined by the ratios of the marginal utilities of those commodities and services.[24]

Given the two behavioural and the two methodological theories four different combinations are logically possible. And indeed the history of economic ideas shows that these four positions were in fact represented by different economists and provide an adequate description of the overall problem situation that prevailed in the early 19th century economic theory. Adam Smith's work, for instance, may be regarded as a combination of a theory of objective evaluations and an individualistic position.[25] To what extent that individualistic position is purely methodological, psychological or ontological is a matter that is to be decided. The work of Marx may be regarded as a combination of methodological or ontological or methodological inductivist essentialism and a theory of objective evaluations. Two more possibilities remain and these are typical of early 19th century economics: the position that consists of methodological individualism and of a theory of subjective evaluations and the positions that represent a combination of methodological inductivist essentialism and a theory of subjectivist evaluations.[26] Whereas the former one, i.e., the position combining individualism and a theory of subjective evaluations is represented by the works of G. Hufeland[27], the latter one is represented by W. Roscher who sought to solve price theoretical problems by retaining a version of Hufeland's theory of subjectivist evaluations but replacing his individualistic approach by an inductivist essentialist one.[28]

Hufeland's work *Neue Grundlegung der Staatswirthschaft* is a response to Smith's *An Inquiry into the Nature and Causes of the Wealth of Nations.* Although Hufeland admires Smith and his analysis in general and his individualistic approach in particular, he criticises Smith's theory of objective evaluations and his price theory as invalid and false and emphasises that this objective value theory may at best be regarded as a special case of the more general theory of subjective evaluations. The arguments he applies in order to demonstrate the falsity of Smith's labour theory are quite similar to those used at present.[29] However, Hufeland's task is not just to demonstrate the falsity of Smith's theory of objective evaluations as a basis for price theory, but rather to improve the prevailing economic theory on the basis of a new price theory. In order to achieve this Hufeland develops a radical theory of subjectivist evaluations and tries to

develop on that basis a new price theory, which, however, is unsuccessful since it lacks the idea of marginal utility and cannot explain how utility considerations determine relative prices. Instead Hufeland develops a theory of reservation prices and considers the special case of a bilateral monopoly in order to explain the process of how prices are determined and scrutinises the different factors that determine the individual reservation prices of individuals who supply and demand.[30] A characteristic feature of Hufeland's approach is that he does not aim at establishing a new empirical basis. He regards facts or processes explicitly as problems that need explanation and emphasises that facts are interesting only because they constitute explanatory problems. According to his view the factual basis that economics commands is sufficiently large and he warns the reader not to expect many new facts, but that new and improved explanations of already well-known facts and processes are given.[31] This – as one might say – anti-inductivist tendency of his work is combined with a methodological individualistic position as well as an ontological one. Although Hufeland considers Smith's value and price theory to be false, he points out that Smith's individualistic approach is most fruitful particularly if combined with a radical theory of subjectivist evaluations. In his view an individualistic position remains unfruitful unless it is combined with a theory of subjectivist evaluations, since only the latter explains social facts and processes in terms of an interplay of individuals' evaluations and knowledge. Although different theories, one being a methodological one, the other a behavioural one, both theories seem to need each other in order to lead to fruitful results.

Hufeland's work appeared in two editions, their difference being monetary theory, which in contrast to the second edition was not written by Hufeland in the first edition. The first edition appeared in 1807 and the second in 1815 and briefly after that Hufeland died.[32] It is interesting to note, however, that Hufeland's theory of subjective evaluations and price theory hardly had any direct impact on the classical English tradition quite in contrast to the further development in Germany. Authors such as Rau[33] and Hermann[34] developed different price theories on the basis of theories of subjective evaluations that emerged somehow or other from Hufeland's ideas. Like Hufeland all these authors were unsuccessful in developing satisfactory price theories but in contrast to him they more or less rejected the idea of explaining the social and institutional framework in which individuals act according to their economic interests on an individualistic basis. In order to explain the coherence of a society they introduce a separate explanatory principle, the "public spirit" that prevents individuals from acting egoistically without limits and introduces altruistic values that hold society together. To them an individualistic explanation of the coherence of society seemed impossible.

The underlying discussion here mixes concepts like egoism and individualism, altruism and collectivism. However, given the open price theoretical problems some authors believed that new methods were needed and in 1843 W. Roscher published a slim volume of lecture notes that had the title *Grundriß zu*

Vorlesungen über die Staatswirthschaft. Nach geschichtlicher Methode[35], that provided an outline of the research programme of the so-called Historical School of German Economics. The epistemological and methodological ideas upon which that research programme was based, had already been published by Roscher in a volume with the title *Leben, Werk und Zeitalter des Thukydides*[36]. In this volume, inspired by the revival of ancient, primarily Aristotelian, ideas through Hegel and Ranke, Roscher develops the position of a methodological inductivist essentialism that regards all social and cultural institutions of a people as emanations of its spirit. Since cultural and social institutions come into existence in the actual historical development the study of their historical development yields genuine scientific knowledge. According to Roscher history is the empirical basis for the theoretical social sciences and the theory of history, and the task of the social sciences is to uncover laws of historical development by the comparison of the development of different peoples. Roscher, however, does not emphasise the essentialist component of his ideas too much, but gives his analysis a naturalistic twist and supports his ideas by an inductivist epistemology[37]. Like all inductivists Roscher holds that genuine new scientific knowledge can be obtained only by inductivist procedures and as a consequence of that view runs into a demarcation problem and the problem of applying inductive methods in the social sciences. In order to propose the correct methodological principles of social science Roscher explains the process of obtaining new knowledge by introducing a passivist psychology of knowledge. According to that theory the human mind is a digestive system which processes the incoming flow of sense data which are obtained by the sense organs, and it is explained that the results of intellectual activities are the product of this process. This passivist psychology is linked to a phylogenetic theory of the development of cognitive faculties, which, according to the intellectual products that result at each stage, distinguishes different stages. Basic ones, where the results are utterances or gestures only; more developed ones, that are characterised by intellectual products such as art and music and even higher ones, where the intellectual results are scientific theories. The products of the highest stage of development are philosophical systems. Whereas the analysis of the first two levels runs in psychological terms, the third one provides a rudimentary methodology of the positive sciences, in particular that of social science, which however proceeds in "naturalistic" i.e. descriptive sociological terms. In his naturalistic account of social scientists Roscher simply translates his passivist psychology into a description of scientific activities. Accordingly, he distinguishes between "historical craftsmen" who collect data and "masters of history" who develop theories and discover the laws of historical development. The historical craftsmen collect the data and facts, which are the raw material or the input from which science starts and which constitute the empirical basis on which the theoretical building is erected. The task of building this theoretical superstructure, the discovery of laws, is fulfilled by the masters of history. They select the relevant data and facts and process them into theories by discovering regularities and similarities.

This naturalist methodology of the social sciences provides the simple inductivist picture of inductivism according to which science starts from unprejudiced observations and builds its theories upon an absolutely certain empirical basis by drawing content-enlarging and truth-preserving inferences from that basis. Roscher's analysis runs, of course, into several difficulties. For instance, the creativity of humans, which raises certain problems at the second development stage of his phylogenetic theory, Roscher explains by introducing "filters" that explain why similar sense data trigger in one case music and in another one perhaps drawings. The problem of being aware of the relevant data is explained by introducing special faculties in terms of a special congeniality between the scientist and his object of research. However, the basic inductivist picture always remains: data, data processing, result.

Since all human intellectual results are according to Roscher the results of inductivist procedures he runs into a demarcation problem, i.e., the problem of how to demarcate empirical sciences from results of the other intellectual activities, such as poetry or music. In order to solve that problem he introduces a rudimentary correspondence theory of truth, according to which theories have to correspond to facts. Roscher considers the introduction of this theory as absolutely necessary because for him creativity, although to a certain extent explicable, always remains a bit of a mystery and only by the acceptance of that correspondence theory is it certain that the edifice of science consists of theories whose truth is proven and certain. Other results of intellectual activities do not have to meet that standard; in Roscher's view the results of art or music need to correspond to certain laws of aesthetical sentiment and philosophical systems need to be logically consistent only.

One of the methodological problems that remains to be solved for Roscher is the problem of applying inductive methods in the social sciences. Roscher shares the view that the natural universe is characterised by the so-called uniformity of nature, that repeated observations are possible, and that inductive methods are applicable in the natural sciences. But according to him the observation of repeated instances is hardly possible in the social sciences, since its empirical basis is the history of mankind and the result of highly complicated and complex interaction of all sorts of different causes. In his view this complexity of the empirical basis of the social sciences precludes the direct application of inductive methods in the social sciences, though he retains the view that only inductive methods, i.e., truth-preserving and content-enlarging inferences promote the growth of knowledge. In order to solve that problem he suggests substituting the idea of inducing laws directly from some observational basis, by observing and comparing the development of different nations or economies. Following this train of thought he comes to the conclusion that, although it is impossible to detect any regularity simply by observing only one nation, such regularities may be discovered by enlarging the empirical basis. In his view, the discovery of the historical development is possible only if we compare the economic development of all nations and peoples, i.e. their history, and induce these laws from historical facts.

Roscher emphasises, however, that in order to carry out this programme anti-naturalistic methods, particularly so-called holistic methods are needed. Accordingly, he recommends methods that take into account not only the economic development of nations, but also their political, cultural, linguistical and legal changes, since all those factors interact closely in the course of their development. Since nations have their own customs, habits and traditions, which create their unique history and determine their presently existing social structures, Roscher held that the task of economics is not to explain regularities which somehow emerge from the rational behaviour of individuals. Its primary task is instead to analyse the historical development of such "wholes" as nations, economic systems and classes, and to detect their historical laws of development.

Roscher's view that genuine and strictly universal knowledge has the form of historical laws of development is nothing but a specific interpretation of a certain kind of Aristotelian essentialism, which may be called methodological inductivist essentialism. Roscher's research programme triggered all sorts of different historical methods in the social sciences and founded the so-called historical school.[38] Every member of this school developed his own position combining methodological inductivist essentialism and a theory of subjective evaluation. However, these authors were not able to solve the problem of how individual preferences or utility considerations determine relative prices and did not develop a unified price theory on the basis of a theory of subjective evaluations. Some authors sought to solve that problem by suspending the development of theories altogether and by focusing on historical studies only in order to constitute an absolutely certain empirical basis for the social sciences. They thought that after having obtained such a basis, better theories could be inferred from it. Other authors did not share this view and thought that the development of economic theory should proceed together with historical research. And some authors already indicated the idea of marginal utility though they did not apply it systematically in order to develop a unified price theory.

It was not until Carl Menger published his *Grundsätze der Volkswirtschaftslehre*[39] in 1871 that this problem was solved. Menger reanimated Hufeland's research programme and substituted methodological individualism for the historical school's methodological inductivist essentialism. He explicitly admired the results that German economists had achieved with respect to price theory, though he thought that their essentialist methodology prevented them from solving the remaining open problem of a unified price theory. His *Grundsätze*, however, not only aimed to solve these problems of economic theory, but also attempted to show that only the combination of methodological individualism and a theory of subjective evaluations provided that fruitful approach that was needed for the progress of economic theory. In order to show this Menger not only stuck to economic theory but also criticised Smith's account of methodological individualism, pointing out that his naturalistic assumption about the individual's faculty to trade and barter does not solve the problem of explaining exchange.

It seems that Menger's contribution to economics was at first well received by his German colleagues. Roscher mentions him in his *History of Economic Thought* [40], though Menger must have noticed that his methodological analysis was not mentioned. Obviously his intention to show the fruitful approach of combining methodological individualism and a theory of subjective evaluations and substituting this combination for the one that combined a theory of subjective evaluations and methodological inductivist essentialism was not even discussed. Twelve years after his *Grundsätze* appeared Menger published a second book, *Untersuchungen zur Methode der Socialwissenschaften und der Politischen Oekonomie insbesondere* [41] in which he delivered a devastating criticism of the methodological inductivist essentialism of the German Historical School and again emphasised the fruitfulness of the combination of methodological individualism and a theory of subjective evaluations. But whereas in the *Grundsätze* he attempted to show the fruitfulness and importance of that approach specifically for the progress of economic theory, he now tried to demonstrate its fruitfulness for the social sciences in general.

In view of the methodological inductivist essentialism held by the authors of the German historical school, Menger discusses several problems connected with induction. He shows that satisfactory explanations in the theoretical social sciences need to propose social laws that are strictly universal and empirical. He also shares the view of many historical economists that such laws may be induced from observations or introspection by inductive methods. However, he rejects the notion of historical laws of development and also the idea of some historical economists to provide an absolutely certain empirical basis for logical reasons. He points out that even singular observational statements have to use universals and that therefore the historical school's aim to avoid abstractions in order to improve the realisticness of theories is logically not possible. He also rejects the historical school's idea that methodological problems can be solved by referring to ontological assertions. Like the authors of the historical school he is an inductivist, however, realising in contrast for instance to Roscher or Schmoller, that the logical objection against inductive inferences needs resolution. In order to do so he introduces a kind of induction principle which he defends on apriori and on empirical grounds. After having shown that we have the choice of either rejecting the fundamental thesis of empiricism, that all empirical statements are decided by experience, or the claim that theories are strictly universal, he proceeds to discuss methodological individualism. He contrasts this position with that of an organic theory of the state and maintains that the former solves basically all the problems that the latter unsuccessfully attempts to.

4. SOME AFTERTHOUGHT

The positions of Menger and Roscher are only two but they very clearly exemplify the central role that the problem of induction and other problems closely related to it, such as the applicability of inductive methods in the social sciences and the demarcation of social science from other areas of investigation, played in the social sciences' epistemological and methodological discussion. In particular it seems to be most fruitful to regard the different epistemological and methodological positions held in the social sciences as attempts to solve – as Popper said – the two fundamental problems of epistemology. And indeed it is quite surprising – or rather not, since they *are* the two fundamental problems of epistemology – how these different positions conform to the different types of their solutions, that Popper explains in his *Die beiden Grundprobleme der Erkenntnistheorie*. Mill, Jevons, Marshall, Menger, Mises, Hayek, Roscher, Knies Hildebrand, Schmoller, Schumpeter and many other authors tried to find a solution for the problem of induction or at least for its applicability in the social sciences. Apart from Menger, however, many authors tried to support their methodological claims by ontological assertions, particularly referring to the so-called "uniformity of nature" as a precondition for the application of induction. This argument, however, cannot support any methodological claim as Menger had already indicated, and as Popper and Miller have shown. Though it is invalid it is still much in use today. But induction is, as Popper says, a myth and myths should be expelled to the realm of fairytales, although some of them have so much influence that they are hardly given up, even if logic advises doing so.

NOTES

1. K.R. Popper, *The Poverty of Historicism*, Routledge and Kegan Paul, London 1957.
2. F. Kaufmann, *Methodenlehre der Sozialwissenschaften*, Pringer, Wien New York, 1999.
3. K.R. Popper, *The Logic of Scientific Discovery*, Hutchinson, London 1959. K.R. Popper, *Die beiden Grundprobleme der Erkenntnistheorie*, J.C.B. Mohr (Paul Siebeck) Tübingen 1979.
4. Cf. M. Schlick, *General Theory of Knowledge*, Library of Exact Philosophy 11, Springer, Wien New York 1974. On p. 386 Schlick for instance writes: "What powers extend our knowledge of past and present facts over to the remote and the future? That they are not the powers of thought, of reason, we know from previous considerations. The inferences of the understanding are by their very nature analytic; they only develop particular truths from the general truths in which they are already contained. Nor can thinking provide more than this. It only gives order and connection to knowledge already gained by means of deductive inferences (§15); it does not create any knowledge (§§39, 40). On the other hand, induction *does* yield knowledge in the highest degree; it is through induction that we obtain the content of all our sciences of reality." ("§"refer to "§" of his *General Theory of Knowledge*.
5. D. Hume, *Treatise of Human Nature*, Scientia Verlag Aalen 1964;

6. In his *Objective Knowledge*, Oxford, 1972, p. 5, Popper quotes a passage from B. Russell's *A History of Western Philosophy*, in which Russell describes this dramatic situation. cf. B. Russell, *A History of Western Philosophy*, George Allen & Unwin Ltd, London 1961, p. 646.
7. K.R. Popper, *The Logic of ScientificDiscovery*, Hutchinson, London 1959.
8. K.R. Popper, *Die beiden Grundprobleme der Erkenntnistheorie*, J.C.B. Mohr (Paul Siebeck), Tübingen 1979.
9. Schlick, for instance, after having explained that only induction yields knowledge of the world we live in continues after the passage quoted in ftn 4 of this paper as follows: "Yet induction cannot be explained by experience any more than by thought, since it extends our knowledge to cases of which we have as yet *no* experience, to wit, cases that are temporally and spatially distant.
   I believe that there is only one answer to the question of the actual origin of inductively obtained propositions and that philosophy has long been in possession of it, thanks above all to Hume.
   If we ask to which human capability we owe knowledge of this sort, we can find no other psychological basis than *habituation*. And habituation in turn rests entirely on the process of association. The image of burning has firmly tied to the combination of the ideas of paper and fire, and the image of cold to the appearance of frost. I am equipped by nature with an association mechanism that allows me without further ado to expect the second term as soon as the first appears, assuming that I have experienced the union of the two often enough..." (p. 386) " .... Had such thousand-fold experience *not* taken place, we could not in fact draw that inductive inference" (p. 387).
10. For a detailed reconstruction of anti- and pronaturalistic positions cf. K. R. Popper, *The Poverty of Historicism*, Routledge and Kegan Paul, London 1961, pp. 5-55.
11. J.S. Mill, *A System of Logic Ratiocinative and Inductive*, book VI, chap. III ff. University of Toronto Press, Toronto Buffalo, 1974.
12. Cf. F. Meinecke, *Zur Theorie und Philosophie der Geschichte*, (Hrsg. E. Kessel) in: F. Meinecke, Werke Bd. IV, (Hrsg. H. Herzfeld, C. Hinrichs, W. Hofer), K.F. Koehler Verlag, Stuttgart 1965, p. 5ff. J.S. Mill, *A System of Logic Ratiocinative and Inductive*, book VI, chap. IV, § 3., University of Toronto Press, Toronto Buffalo, 1974.
13. K.R. Popper, *The Open Society and its Enemies*, Vol. II, Routledge and Kegan Paul, London 1966, p. 7. K.R. Popper, *The Poverty of Historicism*, Routledge and Kegan Paul, London 1961, pp. 26ff.
14. Cf. for instance F.A. Hayek, 'Individualism: True and False', in: F.A. Hayek, *Individualism and Economic Order*, Routledge and Kegan Paul, London, 1949. F.A. Hayek, 'Degrees of Explanation', in: F.A. Hayek, *Studies in Philosophy and Economics*, Routledge and Kegan Paul, London, 1967.
15. Cf. for instance K.R. Popper, 'Models, Instruments, and Truth', in: K.R. Popper, *The Myth of the Framework*, (ed. M.A. Notturno), Routledge, London and New York 1994, pp. 154ff.
16. J.S. Mill, *A System of Logic Ratiocinative and Inductive*, book VI, University of Toronto Press, Toronto Buffalo, 1974.
17. With respect to the distinction of methodological and metaphysical positions cf.: K.R. Popper, *Conjectures and Refutations*, Routledge and Kegan Paul, London 1963, p. 184. K.R. Popper, *The Logic of Scientific Discovery*, Hutchinson, London 1962, p. 55. K.R. Popper, *The Open Universe, an Argument for Indeterminism*, (ed. W.W. Bartley, III) Rowman and Littlefield, Totowa, New Jersey 1982. J.W.N. Watkins, 'The Unity of Popper's Thought', in: P.A. Schilpp (ed.), *The Philosophy of Karl Popper*, The Library of Living Philosophers, Vol. XIV, book I, Open Court, La Salle, Illinois 1974, p. 371.
18. Menger writes: "... among the methodologists no view is more widespread than the one that in certain realms of the world of phenomena, and predominantly in that of nature, strict types and typical relationships can be observed; in others on the contrary, and especially in that of social phenomena only those of lesser strictness can. In other words, 'laws of nature' can be observed only in the first realm; only 'empirical laws', on the contrary, can be observed in the latter.
   However, this assertion of different representatives of the Historical School, "... so frequently found in the general theory of knowledge, will prove erroneous in the sequel. [For it is due to the error that] ... which on close examination turns out to be the result of different *orientations*

of theoretical research in the individual realms of the world of phenomena is construed as the result of the differing natures of the phenomena."
C. Menger, *Untersuchungen über die Methode der Socialwissenschaften und der Politischen Oekonomie insbesondere*, Leipzig 1883. Translated and edited by L. Schneider, *Investigations into the Method of the Social Sciences with Special Referenceto Economics*, New York, 1985, pp. 50, 51.

19. K.R. Popper, *Realism and the Aim of Science*, (ed. W.W. Bartley, III), Hutchinson, London 1983, pp. 71ff., especially p. 76.
20. Cf. D. Miller, *Critical Rationalism, A Restatement and Defence*, Open Court, Illinois 1994, pp. 25ff.
21. Cf. Karl Knies, *Politische Oekonomie nach geschichtlicher Methode*, Braunschweig 1853.
22. Cf. for instance: W. Roscher, *Grundriß zu Vorlesungen über Staatswirthschaft. Nach geschichtlicher Methode*, Göttingen 1843. W. Roscher, *Grundlagen der Nationalökonomie*, Stuttgart 1886. W. Roscher, *Geschichte der Nationaloekonomik in Deutschland*, München 1874.
23. Cf. M. Blaug, *Economic Theory in Retrospect*, Cambridge University Press, Cambridge 1978, chaps 2 and 4.
24. Cf. M. Blaug, *Economic Theory in Retrospect*, Cambridge University Press, Cambridge 1978, chaps 8 and 9.
25. A. Smith, *An Inquiry into the Nature and Causes of the Wealth of Nations*, Oxford 1979.
26. K. Marx, *Das Kapital*, Otto Meißners Verlag, Hamburg 1922.
27. G. Hufeland, *Neue Grundlegung der Staatswirthschaftskunst*, Gießen und Wetzlar, 1807.
28. W. Roscher, *Grundriß zu Vorlesungen über Staatswirthschaft. Nach geschichtlicher Methode*, Dietrich, Göttingen 1843. W. Roscher, *Grundlagen der Nationalökonomie*, Stuttgart 1854.
29. G. Hufeland, *Neue Grundlegung der Staatswirthschaftskunst*, Gießen und Wetzlar, 1807.
30. G. Hufeland, *Neue Grundlegung der Staatswirthschaftskunst*, Gießen und Wetzlar, 1807, pp. 121-160.
31. G. Hufeland, *Neue Grundlegung der Staatswirthschaftskunst*, Gießen und Wetzlar, 1807, preface, pp. 21, 22. [The preface of Hufeland's 1807 does not provide page numbers; these page numbers were assigned by me; K.M.]
32. For a more detailed study of Hufeland's methodological position and contribution to economics cf.: K. Milford, 'Hufeland als Vorläufer von Menger und Hayek', in: *Wert, Meinung, Bedeutung: Die Tradition der subjektiven Wertlehre in der deutschen Nationalökonomie*, (hrsg. B.P. Priddat), Metropolis, Marburg 1997.
33. K.H. Rau, *Grundsätze der Volkswirthschaftslehre*, Winter, Heidelberg, 1826.
34. F.B. Hermann, *Staatswirthschaftliche Untersuchungen*, München 1832.
35. W. Roscher, *Grundriß zu Vorlesungen über Staatswirthschaft. Nach geschichtlicher Methode*, Dietrich, Göttingen 1843.
36. W. Roscher, *Leben, Werk und Zeitalter des Thukydides*, Göttingen 1842.
37. For a detailed analysis of Roscher (1842) cf: K. Milford, 'Roscher's epistemological and methodological position and its importance for the Methodenstreit', in *Journal of Economic Studies*, Vol. 22, Nr. 3/4/5; MCB University Press Ltd., 1995.
With respect to the 'natural twist' of Roscher's methodology cf: M. Weber, 'Roscher und Knies und die logischen Probleme der historischen Nationalökonomie, in: M. Weber, *Gesammelte Aufsätze zur Wissenschaftslehre*, Tübingen 1985.
38. Cf. the survey of the German Historical School of Economics in: J. A. Schumpeter, *History of Economic Analysis*, Allen and Unwin, London, 1986, pp. 808ff.
39. C. Menger, *Grundsätze der Volkswirtschaftslehre*, Wien 1871.
40. W. Roscher, *Geschichte der National-Oekonomik in Deutschland*, Oldenburg, München, 1874.
41. C. Menger, *Untersuchungen zur Methode der Socialwissenschaften und der Politischen Oekonomie insbesondere*, Duncker und Humblot, Leipzig 1883. For a more detailed analysis of Menger's position cf: K. Milford, *Zu den Lösungsversuchen des Insduktionsproblems und des Abgrenzungsproblems bei Carl Menger*, Verlag der österreichischen Akademie der Wissenschaften, Wien 1989. K. Milford, 'Menger's Methodology', in: *History of Political Economy*, Vol. 22a, (ed. B. Caldwell), Durham and London 1990.

REFERENCES

Blaug, M. (1978), *Economic Theory in Retrospect*, Cambridge University Press, Cambridge.

Hayek, F.A. (1949), 'Individualism: True and False', in: F.A. Hayek, *Individualism and Economic Order*, Routledge and Kegan Paul, London.

Hayek F.A.(1967), 'Degrees of Explanation', in: F.A. Hayek, *Studies in Philosophy and Economics*, Routledge and Kegan Paul, London.

Hermann, B.W. (1832), *Staatswirthschaftliche Untersuchungen*, München.

Hufeland, G. (1807), *Neue Grundlegung der Staatswirthschaftskunst*, Gießen und Wetzlar.

Hume, D. (1964), *Treatise of Human Nature*, Scientia Verlag Aalen;

Kaufmann, F. (1999), *Methodenlehre der Sozialwissenschaften*, Springer, Wien New York.

Knies, K. (1853), *Politische Oekonomie nach geschichtlicher Methode*, Braunschweig.

Marx, K. (1922), *Das Kapital*, Otto Meißners Verlag, Hamburg.

Meinecke, F. (1965), *Zur Theorie und Philosophie der Geschichte*, (Hrsg. E. Kessel) in: F. Meinecke, Werke Bd. IV, (Hrsg. H. Herzfeld, C. Hinrichs, W. Hofer), K.F. Koehler Verlag, Stuttgart.

Menger, C. (1871), *Grundsätze der Volkswirtschaftslehre,* Wien 1871.

Menger, C. (1883/1985), *Untersuchungen über die Methode der Socialwissenschaften und der Politischen Oekonomie insbesondere*, Leipzig. Translated and edited by L. Schneider, *Investigations into the Method of the Social Sciences with Special Reference to Economics*, New York.

Milford, K. (1989), *Zu den Lösungsversuchen des Insduktionsproblems und des Abgrenzungsproblems bei Carl Menger*, Verlag der österreichischen Akademie der Wissenschaften, Wien.

Milford, K. (1995), 'Roscher's epistemological and methodological position and its importance for the Methodenstreit', in *Journal of Economic Studies*, Vol. 22, Nr. 3/4/5; MCB University Press Ltd.

Milford, K. (1997), 'Hufeland als Vorläufer von Menger und Hayek', in: *Wert, Meinung, Bedeutung: Die Tradition der subjektiven Wertlehre in der deutschen Nationalökonomie*, (hrsg. B. P. Priddat), Metropolis, Marburg.

Milford, K. (1990) 'Menger's Methodology', in: *History of Political Economy*, Vol. 22a, (ed. B. Caldwell), Durham.

Mill, J.S. (1974), *A System of Logic Ratiocinative and Inductive,* University of Toronto Press, Toronto Buffalo.

Miller, D. (1994), *Critical Rationalism, A Restatement and Defence*, Open Court, Illinois.

Popper, K.R. (1957), *The Poverty of Historicism*, Routledge and Kegan Paul, London.

Popper, K.R. (1959), *The Logic of Scientific Discovery*, Hutchinson, London.

Popper, K.R. (1963), *Conjectures and Refutations*, Routledge and Kegan Paul, London.

Popper, K.R. (1966), *The Open Society and its Enemies*, Routledge and Kegan Paul, London.

Popper, K.R. (1972), *Objective Knowledge*, Oxford.

Popper, K.R. (1979), *Die beiden Grundprobleme der Erkenntnistheorie*, J.C.B. Mohr (Paul Siebeck) Tübingen.

Popper, K.R. (1982), *The Open Universe, an Argument for Indeterminism*, (ed. W.W. Bartley, III) Rowman and Littlefield, Totowa, New Jersey 1982.

Popper, K.R. (1983), *Realism and the Aim of Science*, (ed. W. W. Bartley, III), Hutchinson, London.

Popper, K.R. (1994), 'Models, Instruments, and Truth', in: K.R. Popper, *The Myth of the Framework*, (ed. M.A. Notturno), Routledge, London and New York.

Rau, K.H. (1826), *Grundsätze der Volkswirthschaftslehre*, Winter, Heidelberg.

Roscher, W. (1842), *Leben, Werk und Zeitalter des Thukydides*, Göttingen.

Roscher, W. (1843), *Grundriß zu Vorlesungen über Staatswirthschaft. Nach geschichtlicher Methode*, Göttingen.

Roscher, W. (1874), *Geschichte der Nationaloekonomik in Deutschland*, München.

Roscher, W. (1886), *Grundlagen der Nationalökonomie*, Stuttgart.

Russell, B. (1961), *A History of Western Philosophy*, George Allen & Unwin Ltd, London.

Schlick, M. (1974), *General Theory of Knowledge*, Library of Exact Philosophy 11, Springer, Wien New York.

Schumpeter, J.A. (1986), *History of Economic Analysis*, Allen and Unwin, London.
Smith, A. (1979), *An Inquiry into the Nature and Causes of the Wealth of Nations*, Oxford.
Watkins, J.W.N. (1974), 'The Unity of Popper's Thought', in: P.A. Schilpp (ed.), *The Philosophy of Karl Popper*, The Library of Living Philosophers, Vol. XIV, book I, Open Court, La Salle, Illinois.
Weber, M. (1988), 'Roscher und Knies und die logischen Probleme der historischen Nationalökonomie', in: M. Weber, *Gesammelte Aufsätze zur Wissenschaftslehre*, Tübingen.

Dept. of Economics
University of Vienna
Hohenstaufengasse 9
A-1010 Vienna
Austria
karl.MILFORD@univie.ac.at

DAVID MILLER

THE UNIFORMITY OF NATURE: WHAT PURPOSE DOES IT SERVE?

Comments on Karl Milford's "Inductivism in 19th Century German Economics"

0.

In the third section of his paper 'Inductivism in 19th Century Economics' Karl Milford presents us with a useful résumé of the impact of methodological considerations, especially amongst economists in the German-speaking world, on more properly economic problems, especially the development of price theory. It is interesting to learn from his report how diverse were the various approaches, both axiologically and methodologically. Each combination of objectivism or subjectivism with individualism or collectivism had distinguished representatives: the objectivism with regard to values of Smith and Marx contradicted the subjectivism of Hufeland and Roscher, while the individualism of Smith and Hufeland contradicted the collectivism of Marx and Roscher (p. 10). As Milford tells the story (p. 285), it was only with publication of Menger's *Grundsätze* in 1871, and his *Untersuchungen* in 1883, that the combination of subjectivism and individualism espoused by Hufeland finally attained a dominant position. The success of Menger's approach seems to be attributable on the one side to the persistent inability of objectivist theories of values and prices to explain adequately the activity of the market; and on the methodological side, to his devastating attack on some aspects of the naively unaffected inductivism underlying most collectivist thinking (for example, the thinking of Roscher). In fact Menger too was an inductivist who, in response to Hume's attack on all modes of content-increasing inference, introduced and defended an a priori principle of induction. In the light of such an entrenched tradition of explicitly inductivist thinking, it is puzzling that the widely praised *Cambridge Dictionary of Philosophy* prefers to characterize inductivism as 'a philosophy of science invented by [Karl] Popper and P.K. Feyerabend as a foil for their own views' (p. 426). 'According to inductivism,' the brief article continues (and concludes) 'a unique apriori inductive logic enables one to construct an algorithm that will compute from any input of data the best scientific theory accounting for the data.' This terminology leaves us at a loss for a name for the most widespread methodological prejudice in the social sciences.

F. Stadler (ed.), Induction and Deduction in the Sciences, 293–297.
© 2004 *David Miller.*

Milford tells us that pro-naturalistic and anti-naturalistic philosophers of society (p. 274)

> disagree with respect to the question of the possibility of making repeated observations for the social sciences. Whereas they agree that this possibility exists to a very high degree for the natural sciences, because the 'natural universe' is ontologically characterised by the so called 'uniformity of nature', they emphasise that the 'social universe' is much more liable to change. From this it is concluded that the method of induction cannot be applied in the social sciences as rigorously as in the natural sciences. However, opinions diverge with respect to what degree this possibility exists.

Later he adds that 'Roscher shares the view that the natural universe is characterised by the so-called uniformity of nature, that repeated observations are possible, and that inductive methods are applicable in the natural sciences ... [but] the observation of repeated instances is hardly possible in the social sciences' (p. 284). Plainly read, the principle of the uniformity of nature here appears as a statement about the existence of observable regularities rather than of laws of nature. It is a low-level factual statement, that can fairly be regarded as a logical consequence of any particular statement of observable regularity. Neither it nor its denial says anything about the existence of more delitescent regularities; and indeed many inductivists have thought that there do exist such undisclosed regularities in the social world, hidden behind the multifariously variegated appearances, regularities that can perhaps be discovered by some less pedestrian method than simple induction by enumeration. If this is so, then the social sciences are not so very different from the natural sciences. Kepler's laws, for example, do not record observable uniformities, but simple abstract connections that reveal themselves at the empirical level in a bewildering medley of planetary motions. It is amusing how apologists for underdeveloped sciences like to emphasize the complexity of the phenomena for which they have to account, conveniently depreciating the intricacy of those natural phenomena that first yielded to genuine scientific explanation.

It would be presumptuous of me to enter further into the historical issues that Milford so well summarizes. Almost everything that I know about the history of economic thought I have learnt from his writings. But there is perhaps something worth saying from a logician's perspective about the role that is played, or imagined to be played, or denied to be played, in science by what are called 'inductive principles' or 'principles of uniformity of nature'. I hope that what I have to say will shed some light on the various methodological positions that have decorated the history of the natural and the social sciences.

1.

The doctrine that we learn from experience through a process of induction by enumeration is not dead by any means. But many philosophers have accepted

that, if we are to account for unobservational and counter-observational aspects of theoretical science, induction (whether in its pure form or disguised as what Milford calls inductivist essentialism) cannot be the whole story. Ever since Whewell simultaneously praised and dispraised Bacon for his methodological insights, the hypothetical method has been widely agreed to play an indispensable role in the discovery of regularities. In these circumstances a principle of uniformity of nature of the kind described above seems completely superfluous; indeed, it has recognisable methodological significance only when it is denied. Yet it has been thought by many that the search for regularities, whether or not guided by hypotheses, would be pointless if there were not indeed regularities (observational or theoretical) to be searched for. It has been felt, that is, that scientific research presupposes some principle of uniformity of nature, and is therefore at least covertly inductivist. I have explained elsewhere (1994, § 2.2*a*) why this is a mistake. Science does not presuppose uniformity, it proposes it, and then attempts to dispose of its proposals. In order to search for something, it can be necessary only in an irrelevant psychological sense to presuppose that that something, or things like that something exist. Whether it is necessary is strongly to be doubted. To be sure, it would be unwise to presuppose that the thing searched for does not exist (though not to hope that it does not). Mathematicians checking proofs for mistakes, and proof readers checking proofs for errors, neither presuppose nor preclude the existence of what they are looking for.

The further idea that the postulation of some inductive principle can provide a means of warranting or justifying the truth of those universal empirical generalizations that survive investigation is really too ill considered to merit further discussion (*op.cit.*, § 2.3). Hume made that clear enough. If Menger, who was aware of the logical objections to induction, saw fit to introduce a priori a principle of induction, it was not in the vain hope that empirical hypotheses rather than exact laws involving ideal types, would by this means be justified. On this point see Milford (1990), § V.

2.

It should be an obvious conclusion to draw from these remarks that a principle of uniformity is entirely superfluous in the non-justificationist philosophy of falsificationism, the methodology of conjectures and refutations. After all, in § 79 of (1934/1959), Popper explained clearly the manner in which he sought to substitute a fallible methodological rule ('Natural laws should be spatio-temporally invariant') for an unfalsifiable metaphysical principle ('Nature is uniform'). The rule is fallible in the sense that the conjectures that we propose may be falsified one after another, so that we are never in a position to say correctly that we have reached, however tentatively, a law of nature. Of course, nothing justifies the methodological rule, which is understood to be a convention (*ibid.*, § 11).

There have been some recent attempts, however, to upset this conclusion by means of the following argument. (Jüri Eintalu tells me that an anticipation is to be found in Reichenbach (1952).) A principle of uniformity is required in science not as a positive tool (for justification) but as a negative tool (for elimination). Without it, there is no way of excluding from science a host of unusual hypotheses that everyone agrees are highly undesirable. This class of mischief-makers includes not only hypotheses of the kind made famous by Goodman (1955), but also hypotheses such as 'Newton's law of gravitation $F = Gm_1m_2/d^2$ holds for all phenomena except laboratory phenomena where 5lb of gold dust is sprinkled on the laboratory floor, in which case $F = Gm_1m_2/d^3$' (Maxwell 1998, p. 52). Such hypotheses, especially those of the second type, are not excluded by the methodological injunction to propose only falsifiable conjectures that are spatio-temporarily invariant. Indeed they are easily falsifiable, should anyone bother to test them. But, it is argued, no one actually will bother to test them, and on the principle (the principle of empiricism) that hypotheses should be rejected only on empirical grounds, they therefore remain within science. Since there is no limit to the supply of such tiresome variants of respectable theories, falsificationism is mistaken if it supposes that it makes any headway in solving the problem of induction. That problem can be solved only by postulating a principle of comprehensibility and unity that is markedly richer and more elaborate than any principle of uniformity entertained previously (Maxwell *op.cit.*, Chapters 1-3).

I think that this argument is mistaken (Miller 2002, § 4) and that, despite its claim to be a sympathetic but necessary adjustment to falsificationism, it is no less justificationist than its ancestors were. What is represented here as in need of empirical justification is of course not a scientific hypothesis (it is admitted that all hypotheses remain for ever conjectural), or even its acceptance into science for testing, but its exclusion prior to testing. We may ask, however, why any such exclusion is in need of empirical justification. If the answer is that, according to the principle of empiricism, 'in science, only observation and experiment may decide upon the *acceptance or rejection* of scientific statements, including laws and theories' (Popper 1963, Chapter 1, § IX), and this is understood to mean 'in science, only observation and experiment may decide upon *whether or not* scientific statements, including laws and theories, are accepted or rejected', then we must discard the principle of empiricism. Indeed, in this form it leads to an infinite regress, since no empirical argument (argument from observation and experiment) can be presented for or against any statement A (including the statement that a hypothesis h ought to be accepted or excluded) until that statement A has been accepted. To have an argument to accept or to exclude h, therefore, we must already have accepted A. We shall plainly be able to act more efficiently if we are prepared to accept or to exclude h without any argument. Excluding a hypothesis is after all not a way of rejecting it. What the principle of empiricism, in the revised form that is crucial to falsificationism, asserts is that no hypothesis that has been accepted may be rejected except on the basis of observation and experiment (compare Popper 1959, § 11, which admits

also the possibility that a hypothesis may be replaced by one that is better testable). Rejecting a hypothesis after all is a way of excluding it.

Summarizing his criticism of Carnap's attempt to apply considerations of probability to the problem of induction, Popper wrote (1963, Chapter 11, § 6):

> I believe that what we have observed here shows that there is a third way of violating the principle of empiricism. We have seen how it can be violated by constructing a theory of knowledge which cannot do without a principle of induction – a principle that tells us in effect that the world is (or very probably is) a place in which men can learn from experience; and that it will remain (or very probably remain) so in future.

Of course, if the world is not like this then we shall not succeed in learning from experience. But to suppose that science presupposes that its methods will be successful is to suppose too much (Miller 1994, § 2.2*a*). I continue to maintain that there is no need anywhere in science for a principle so vague and at the same time so devoid of purpose.

REFERENCES

Goodman, N. (1955). *Fact, Fiction, and Forecast.* Cambridge MA: Harvard University Press.

Maxwell, A.N. (1998). *The Comprehensibility of the Universe.* Oxford: Clarendon Press.

Milford, K.M. (1990). 'Menger's Methodology'. In B. Caldwell, editor, *Carl Menger and his Legacy in Economics* (Supplement to *History of Political Economy* **22**), pp. 215-261. Durham and London: Duke University Press.

Miller, D.W. (1994). *Critical Rationalism. A Restatement & Defence.* Chicago & La Salle: Open Court Publishing Company.

Miller, D.W. (2002). 'Induction: a Problem Solved'. In J.M. Böhm, H. Holweg, & C. Hoock, editors, *Karl Poppers kritischer Rationalismus heute*, pp. 81-106. Tübingen: Mohr Siebeck.

Popper, K.R. (1934). *Logik der Forschung.* Vienna: Julius Springer Verlag. 10th edition, 1994. Tübingen: J. C. B. Mohr (Paul Siebeck). English edition, Popper (1959).

Popper, K.R. (1959). *The Logic of Scientific Discovery.* London: Hutchinson Publishing Company. Expanded English translation of Popper (1934).

Popper, K.R. (1963). *Conjectures & Refutations.* London: Routledge.

Reichenbach, H. (1952). *Experience and Prediction.* Chicago: University of Chicago Press.

Department of Philosophy
University of Warwick
COVENTRY CV4 7AL
UK
d.w.miller@warwick.ac.uk

SYBILLA NIKOLOW

PLANNING, DEMOCRATIZATION AND POPULARIZATION WITH ISOTYPE, CA. 1945: A STUDY OF OTTO NEURATH'S PICTORIAL STATISTICS WITH THE EXAMPLE OF BILSTON, ENGLAND

INTRODUCTION

In Otto Neurath's pictorial statistics ("Bildstatistik"), statistics communicate not through numbers or graphical representations but through larger and smaller numbers of symbols. We are dealing with an attempt to create a visual, transparent and universal language for the global communication of information and arguments. Here, I want to further explore Neurath's pioneering work as a visual pedagogue and as an expert on the popularization of knowledge on the example of how his *International System of Picture Education* (ISOTYPE) was used in an exhibition on new housing in Bilston, a town in the Midland Counties of England. I use this case in order to expound the design and practice of his picture-statistical method in the context of contemporary scientific, artistic, political and pedagogical culture.

Neurath's statistics grew out of the so-called "Vienna Method" which he developed in collaboration with Gerd Arntz, a graphic artist from the Rhineland, in the context of the workers' education movement of the interwar years. During his emigration to the Netherlands this method received the international name of ISOTYPE and became expanded and transformed into a universal pictorial language.[1] When Rotterdam was bombed by the Germans on 10 May 1940, Neurath fled from The Hague, this time only with his closest collaborator of long standing and later third wife, Marie Reidemeister. Following his release from internment in Britain in February 1941, the two of them founded the ISOTYPE Institute in Oxford and continued their work on pictorial statistics. They published their own work, illustrated the publications of others and contributed to documentary films and exhibitions.

Neurath returned to his role as scientific advisor, a function familiar to him from his days in Vienna, only shortly before the end of the Second World War. Back in Vienna he had founded the Museum for Society and Economy ("Gesellschafts- und Wirtschaftsmuseum") and he had developed the method of pictorial statistics in order to win the support of the Viennese population for reforms

F. Stadler (ed.), Induction and Deduction in the Sciences, 299–329.
© 2004 *Kluwer Academic Publishers.*

envisaged by the socialist municipal administration. In Bilston, however, he emphasized vis-à-vis the local authorities that the primary goal of educational campaigns now consisted in motivating those concerned to contribute actively in reforms. In this respect Neurath's ideas noticeably stood out from those of other experts and reformers who likewise were active on the interface between planning and public education. Towards the end of this paper I will return to his approach to the popularization of knowledge, which he formulated in a manuscript on *Visual Education. Humanisation versus Popularisation* 1945.[2] This he wrote when he started working on Bilston. Through a closer look at this case in the context of the broad debate about the future, as it then was conducted in Great Britain, we stand to gain a deeper insight into Neurath's proposals for a democratic postwar society.

Neurath was invited to Bilston in July 1945, when the authorities there began to plan the relocation of slum-dwellers to a new housing estate on the edge of the town. In his report Neurath recommended that the citizens concerned should be informed about the plans in advance and that they should be involved as much as possible (Neurath and Cohen 1973, 78). As the instrument of a form of communication that "can impart information while giving pleasure" (ibid.), he suggested an exhibition, for he hoped that the citizens would not oppose the plan if it was successfully communicated to them that the move would improve their lives. Five months later, in December 1945, Neurath died unexpectedly. Less than a year later the exhibition took place all the same. Marie Neurath, who now ran the ISOTYPE Institute on her own, supplied it with pictorial charts a selection of which she published in the following year (Neurath 1947).

How were the plans and the problems of Bilston translated into the language of pictorial statistics? How were pictures made to take the place of Neurath, the municipal administration and the architects in talking to the citizens? Was it up to the Neuraths alone to determine what was said in the exhibition, and how it was said, or were they forced to compromise? When, with whom and in what respect? Was the exhibition, or the project, a success and if so, should we attribute this to the pictorial statistics? In order to answer these questions we must look beyond the published accounts of this exhibition, all of which appeared later, and to contemporary press coverage and the correspondence between Marie Neurath and A.V. Williams, the mayor of Bilston, which is preserved among the papers of the ISOTYPE Institute.[3] The detailed analysis offered here of the processes of communication and labour involved in Neurath's pictorial statistics will allow us to assess his theorizing and the subsequent recollections of Marie Neurath in their respective historic context (Neurath 1947; Neurath and Cohen 1973, 75-78). It will also enable us to reach more far-reaching conclusions about Bilston than permitted by the literature to date (Faludi 1996, Kräutler 1996, 197).

THE PROBLEMS OF BILSTON AND NEURATH'S SUGGESTION FOR THEIR SOLUTION

Bilston is situated to the west of Birmingham in the Black Country. The industrial revolution left distinctive marks on the local landscape and living conditions (Williams 1947, 251).[4] Iron production was followed by the steel industry, whose works spread through the town because the hinterland already was entirely exhausted (Bilston 1998, 10). This helped prevent the economic decline of the region yet living conditions, particularly those of the workers, became ever more unbearable. The inhabitants of Bilston suffered from the unusually high population density, from the catastrophic sanitary conditions in most of the houses and from excessive air pollution caused by the industry (Williams 1947, 252). In a door-to-door survey which the local authorities had carried out shortly before the end of the Second World War, over a third of all houses were declared "unfit for habitation" (Williams 1947, 252).

Two months after the end of the war and one month after the Labour victory in the national election, the city authorities began to tackle this problem on a large scale, which meant to involve the surrounding countryside in the search for a solution (Neurath and Cohen 1973, 75-78). For Williams the time seemed to have arrived at last when those could be helped who had contributed most to the yield of the industrial revolution yet of whom it was still demanded "to live in conditions which are not merely depressing, dirty and inconvenient, but which actually constitute a serious menace to the health of the parents and the lives of their children" (Williams 1947, 253). Williams was not alone with this take on the situation. As we know from various historical studies, towards the end of the war regional governments in Great Britain increasingly regarded planning – and town planning in particular – as an opportunity to settle the "essential part of the equation" (Gold and Ward 1997, 60) with those portions of the population who had made the greatest sacrifice, first owing to the consequences of the industrialization and then during the Blitz.

At heart Neurath's recommendations consisted in three suggestions: firstly, to get across to the citizens that the realization of the scheme would increase their standard of living; this was to be achieved through the exhibition. Secondly, to study existing social structures, for families were not to be separated unless they so desired. Here Neurath was thinking above all of two statistically underrepresented groups, those over 65 who no longer worked and those living alone, who potentially faced isolation. And thirdly, the administration was to come to the citizens, not the other way round (Neurath and Cohen 1973, 75–78). Indeed, this was a classical demand, made in many town planning schemes of the postwar period.

POSTWAR UTOPIANISM

The town council of Bilston saw Neurath's recommendations realized in their most ideal form in the "Neighbourhood Architecture" of the celebrated British architect Sir Charles Herbert Reilly (1874–1948). Reilly planned around the "Community Principle of Neighbourhood Unit", an architectural doctrine that was first discussed in the United States and which, from the 1940s, became a "major and remarkably enduring planning principle" in Britain (Ward 1994, 96).[5] This principle stipulated that no more than ten thousand people should live together in small estates and that the key social facilities should be accessible without crossing the street. Reilly had worked out a plan according to these specifications for the town of Birkenhead, which had however been rejected by the conservative majority in the town council and which was now taken up by Bilston. Houses were to be arranged around a shared green or playground (subsequently dubbed "Reilly Green" after its instigator), so that only the front doors opened out onto a private garden and the street.[6]

In the "Battle of Birkenhead" (Edelman 1944)[7], so the *Picture Post*, the "Community Principle of Neighbourhood Unit" did not win out. The battle of Bilston had a different outcome, yet this was no longer unusual, for the mood changed towards the end of the war and peaked after the Labour victory in the British House of Commons.[8] Ultimately the quarrel over Birkenhead contributed to the fame of Reilly's plan, which became the subject of a study in social and psychological development in which it was hailed as a "new way of life" with which urban isolation could be fought effectively (Wolfe 1945).

At this time, the postwar utopianism of regional politicians, experts and planners found an important mouthpiece in the press. Headlines like "The Battle of Housing", "The Battle for Health" and "Operation Housing" helped move civil postwar projects like housing and health services into public consciousness. In contrast to military operations the hopes of the democratic public were focussed on transparency and on people's right to have a say in civil planning through local governments. Characteristic of this was, for example, that the *Picture Post* informed its readers in the summer of 1945 not only of the experiences of soldiers in enemy territory, but at the same time inquired critically whether the home nation was sufficiently prepared for the return of its heroes. Middlesborough, a town with problems similar to Bilston, was paraded as a pioneering example. There, citizens had successfully reclaimed their houses, which had been destroyed during the war, and they were involved in the planning for reconstruction on which the *Picture Post* successively reported.[9]

HOW WERE THE SLUM-DWELLERS OF BILSTON ENTICED TO VISIT THE ISOTYPE EXHIBITION?

Fortified by Neurath's advice the Bilston authorities likewise pursued new paths in their attempts to reach those concerned in advance of the realization of their scheme. Unlike the official presentation of the results of the door-to-door survey, which had taken place in the Town Hall,[10] the venue chosen for the exhibition was a little shop in the middle of the slum area for it was known, as the chairman of the municipal health committee announced on the BBC, that "the folk here don't generally go to the large public buildings where such exhibitions are commonly held".[11] This was not enough, however. Neurath argued that "an exhibition which is intended to be a part of a campaign for the immmunization of children against diphtheria could be presented as an exhibition of the wonders of scientific achievements in the field of medical science"; such an exhibition could make "housing and happiness" (Neurath and Cohen 1973, 78) its theme through the back door. He had extensive experience with the comprehensive interpretation of the theme of health.[12]

When the opening approached in November 1946, Marie Neurath and Williams began to discuss what materials should be exhibited in addition to the pictorial statistics. They considered several possibilities. Newspaper articles before and after the opening substantiate that in addition to ISOTYPE-charts a model of a "Reilly Unit" was shown, but that the magnet for the public was another, sensational, object. While, interestingly, this was never mentioned in the recollections of the organizers, the *Birmingham Gazette* reported a week before the opening of a "formicarium", an ant colony that could be inspected in a glass case. The case was positioned in such a way that the colony was visible from the street through the shop window but that "if people want to look at the ants they *must* go inside the exhibition" (Anon. 1946).

It is clear from the correspondence between Marie Neurath and Williams that this exhibit came from Derek Wragge Morley. Williams introduced him as a scientist and sociologist.[13] Caricatures on the social life of ants were also on show.[14] Marie Neurath took care to ensure that the additional objects were not pictorial charts. Referring to Otto as her authority she wrote to Williams expressing the hope "that the exhibits you mention are not charts which might in any way clash with the Isotypes in their method of presentation".[15] She raised no objections against putting an ant society on display, though a critical commentary on a scientific manuscript by Morley which she produced at his request hints at slight misgivings in the spirit of Otto's understanding. Using "Otto's pencil", she admonished Morley only to conclude about ants what was warranted by genuine observation alone.[16]

The organizers hoped to achieve much more with the ant hill than "just" to lure visitors inside the shop. The local press used the headline: "Ants will guide

this town to healthier living", suggesting that this exhibit would be appreciated especially by people "who do not readily grasp the implication of statistics" (Anon. 1946). This is how it may have looked to visitors yet ultimately it was the planners who guided the Bilstonians. Formicarium and caricature were used to show that even tiny insects like ants were happy and content in the "Community Principle of Neighbourhood". In the pamphlet that was available at the entrance, a page was used to transmit the ideas of the planners to the Bilstonians with the aid of the ants. Visitors were expected to learn from the ants that they would survive not as individuals but only in societies: "Why then are ants being used in this exhibitions? Isn't this doing just what we say should not be done? No; we can learn nothing of value by comparing their 'milking' of green flies, their putting them out to 'pasture' on leaves or building sheds over them to protect them, and other man-like activities, with those of man. But there are things we can learn by studying their communities."[17] For only those ant communities would survive in which "the ants all appear to work for the common good of the nest, and the work is shared" (Ibid.). In other words, the "Battle of Bilston" was about winning over slum-dwellers to the idea of community. The architectonic model was used to give a spatial impression of how 500 people could live together as a unit. The ants, on the other hand, had shown beyond doubt that such a way of life was natural and in the pictorial statistics, lastly, one was meant to come face to face in the plans with one's dreams of a better life.

THE PICTORIAL STATISTICS

In the pictorial statistics the insights gained from the glass case about the social life of insects were translated into human proportions. Once again the ISOTYPE was put to the task of presenting the plans of experts in such a way as to connect with the language of the observers and to enable them to recognize the problem under discussion as their own and to comment on it. For the ISOTYPE Institute this meant that the unfathomably bad living conditions of the slum population needed to be transformed into a problem that could be grasped and therefore controlled. ISOTYPE stood for a decoding machine that made it possible to translate such information into clear and countable units of symbols, that is, into a problem of land and people or inhabitants, families, the town, its houses in relative size, and so on. Neurath called this step "transformation" and he had trained Marie Reidemeister (later Neurath) in this skill.[18] From the Viennese period we know that Neurath did not attribute to the translation of numbers into images a lesser significance than he did to the design of the pictorial elements.

1. Bilston – the Houses
2. Bilston – the Land and the People
3. Sizes of Families and Houses They Need
4. Houses Sizes Available and Planned

5. Two Types of Planning
6. Who should get a house first?

---

7. Air Pollution in Bilston
8. How 300 persons are housed in old and new parts of Bilston
9. Deaths in the First Year of Life in Bad and in Good Houses
10. Infectious Diseases Spread in Overcrowded Areas

---

11. Facilities for Leisure
12. Why have we to Plan for Leisure Now?

Figure 1
Title of the Charts, IC (MS 1091 3.2/93-96).

In the exhibition 12 charts were shown, each around 4 foot high and in colour, but unfortunately we have reproductions of them only in black and white.[19] The charts (Figure 1) can be divided into four groups. The first four show the plan, the next two were meant to involve the citizens in the planning process by conceding them a right to have a say in specific aspects of the scheme. The second part of the exhibition dealt with the theme of housing more broadly, first addressing health problems in four charts and then leisure facilities in two further charts. These six charts were meant to persuade those concerned that moving was in their interest. As was common in other postwar plans, present or past situations were juxtaposed with future improvements.

In the first chart ("Bilston – the Houses"), houses were classified according to their inhabitability according to the results of the door-to-door survey and it was specified how many families had no house of their own so far. In the second chart ("Bilston – the Land and the People", Figure 2) the present condition was contrasted with the planned future. This chart showed how the land would be redistributed in its new utilization. The illustration was meant to give the impression that the same number of people would find more room if the fallow land would be released for the building of new houses. From the correspondence it transpires that Williams asked Marie Neurath to insert an excavator so as to communicate that this process had already been started by the authorities.[20]

What these dimensions of the planning would mean in terms of the numbers of houses and families was shown in the next two charts. The third one ("Sizes of Families and Houses They Need") illustrated how many families in Bilston would benefit from the project and how big a house each could expect according to the number of family members, converted into the number of rooms. The fourth chart ("House Sizes Available and Planned", Figure 3) showed what this would mean in detail for the construction works, that is, how many houses would need to be repaired and how many of them would need to be built from scratch if the result just presented were chosen as the basis for calculation. In the first four charts the key quantitative dimensions of the plan were introduced. These consisted of fixed basic parameters, like the size of the land, of the population and

the number of houses. These data derived from the door-to-door survey and they were presented as the incontrovertible facts of rational analysis.

The next two charts showed what Neurath meant when he spoke of ISOTYPE as empowering observers, through the juxtaposition of visual facts, to participate in exchanges with experts. The fifth chart ("Two Types of Planning", Figure 4) showed six building plans and confronted the slum-dwellers with two fundamental decisions: should families of the same age and size live together, as shown on the left, or should families of different age and size live together, as shown on the right? Also: how were the houses to be arranged and who was to share a Reilly-Green? Three options were played out horizontally: rows of houses, separated by streets; a circular arrangement with a shared common in the middle, or each house by itself with individual front and back gardens. The decision was not just between two types of planning, as the title suggested, but between two principles of living: in future, did the families in Bilston wish to live by themselves or in communities?

Here planning is portrayed not as a centralized process from the top down but as a partially incomplete process. The idea was not to push the observer into a decision through the demonstrations of the plan, although it must be admitted that Reilly's proposal appeared as an eye-catcher in the centre of the chart, visitors had been introduced to the Reilly-unit through a model of it and the planners had used the ants to impress upon people the notion that only communal life suited the nature of social beings. Yet patently, the chart suggests that families can choose individually rather than by majority.

The sixth chart (Figure 5), "Who should get a house first?", addressed those concerned directly with a question. Here they were offered to assume the role of the planners and to prioritize themselves the neediness of candidates. The following criteria were offered: past housing quality, age and size of the family and the health of its members. The rules of the game were communicated through examples. More clearly than in the previous image observers were encouraged to pursue a particular solution. For the most points were given to families who hitherto had no home of their own, and whose children or other members were unable to leave the house for any length of time due to ill health. The invitation at the bottom edge of the chart, "What would you suggest?", asked visitors of the exhibition to assent to the solutions that had been shown or to formulate an alternative.

At the end of this first part of the exhibition the social implications of the Reilly plan had been introduced that would make his model more graspable for citizens who wanted to know whether they were entitled to a new house, how many rooms they could expect and how they should picture life in the new estate. Here clarifying the plan meant to feed information about its details step-by-step. This strategy continued in the second part of the ISOTYPE-exhibition, where it was expanded further. The goal here was to sketch out an image of life in the future of which the planners assumed that it would capture the wildest

dreams of those at present living in the slums. The ISOTYPE Institute argued with the generally known positive consequences of modernization.

Neurath had mapped out these possibilities for health and leisure as early as 1939, in his *Modern Man in the Making* (Neurath 1939). Both themes were worked into a convincing argument in favour of his plan. The point was to create the impression that the move brought a healthier and happier life within reach. In Neurath's view these were fundamental human needs. It had to be shown that the health problems from which the slum population suffered in particular (rickets, infant mortality and infectious diseases) were caused solely by bad living conditions. If conditions were improved in the near future, so the visual argument ran, these other threats would disappear.

In the seventh chart ("Air Pollution in Bilston", Figure 6), a map of the town of Bilston suggested that air pollution through industry was worst in the slum areas. The nexus between air pollution and bad health was made in the lower half of the image, where lack of sunshine became linked to rickets. This disease – which also went by the name of "English disease" on the Continent – was very widespread among children in the industrial areas of the Midlands and led to irreversible bone damage. As was customary in health education, sickness was juxtaposed with healthy appearance, generating the impression that if certain rules were followed the onset of disease could be prevented from the start or at least that it could be combatted in the early stages, so long as certain therapeutic measures were followed. The unhealthy past can be seen on the left hand and the healthy future on the right, as if one were reading a text, a principle which Neurath deliberately employed in his visual language. In the eighth chart ("How 3000 Persons are Housed in Old and New Parts of Bilston", the dark world of the past on the left contrasts with the bright future on the right. The houses in the new town are further apart and they are brighter because they are taller and let in more light through their windows. As for the question of how the houses are to be arranged, both types of planning are presented next to each other, while the ninth chart focussed solely on greater distances between the houses.

Here ("Deaths in the First Year of Life in Bad and in Good Houses", Figure 7) the depiction of the future is given even more room in the image; it now takes up more than half the chart. Reference is made to the eighth chart through a miniature version of it in the top part, supplemented further down through symbols for babies and crucifixes to represent infant mortality. From 1939 Neurath had used crucifixes both with and without gravestones to symbolize mortality, without actually aiming for consistency.[21] The nexus between living conditions and infant mortality was taken up again in the tenth chart ("Infectious Diseases Spread in Overcrowded Areas") and expanded into the general theme of infectious disease with the headline "Contact is Dangerous – Isolation of Infectious Person Is Important". As a solution to the problem a third bedroom for a family of five was suggested. Since Neurath became active in health education he used radiating lines to convey that someone suffered from an infectious disease and that therefore a danger of infection for healthy persons was

emanating from this person.[22] The caption left no room for doubts but presented a scientific fact to which the visitors of the exhibition were expected to agree after seeing this image.

The last two charts documented the importance Neurath attributed to leisure. He considered increasing prospects of leisure an important product of modern social development, and one whose effects people would feel in their everyday lives. *Modern Man in the Making* already contained the utopian idea that future mankind would have to work even fewer hours than before and people would be able to spend more time with their families and friends.[23] This historic achievement the ISOTYPE Institute wanted to bring home to the inhabitants of Bilston.

The eleventh chart ("Facilities for Leisure") is an ideal-typical comparison of past and present. War here appeared as a watershed that effectively separated the prewar and postwar periods from one another. As the postwar proposal figures the Reilly-plan, according to which all gardens in the back were to be handed over for a shared green and only those out front remained private. What is shown here is not only that before, fewer activities were possible. It is also implied what can be done with the area that will be commonly used (indeed, this was what had sparked off conservative criticism of the Reilly-plan in Birkenhead) and what kind of new opportunities might arise from a decentralized distribution of central institutions. There are symbols for children's playgrounds, clubs, libraries and institutions for children. All of these, so the implicit claim of the image, would be found in a "Reilly Unit", that is, at close distance. The twelfth chart ("Why have we to Plan for Leisure Now?", Figure 8) discussed the idea that owing to the reduction of working hours and increased leisure over the past 100 years a situation had arisen which itself seemed to require planning. As in *Modern Man in the Making*, the clock was transformed into a pie chart to demonstrate the differences in the hours.[24] Except for reading and visits to the pub, nothing seemed the same as it was before. As in the previous chart, social and communal activities were prominently in view. To these were added visits to concerts, the theatre, swimming pools or dances and it was shown that the communal lawn could be used for playing grounds and football pitches.

Pictorial statistics were used to promise the slum population of Bilston a bright future in new houses on the edge of town and to impress the vision that people's dreams for a better life would all be realized with the plan. In the charts people were encouraged to participate in the improvement of their living conditions according to the ideas of the planners. They were shown ways, moreover, in which they could contribute to the settlement of issues that were not yet resolved, e.g. how social justice was to be achieved in the distribution of houses among families, how houses were to be arranged and how communal areas were to be used. Pictorial statistics communicated to the observer the contents of the plans, not only by simply transporting them into the picture but also through the choice of examples and symbols that had affinities with the lives of those concerned, which they could observe and examine themselves, like the decrease in infant mortality or working hours. ISOTYPE offered several ways to

facilitate communication between experts and laypersons, as in the step-by-step broader interpretation of the topic and in the step-by-step expansion of the visual vocabulary that went with this. The charts first dealt with basic needs like housing and sleeping, which were represented by symbols for house and bed. The exhibition did not end here, however. The content of the planning having been explained, this created the basis for directing the attention of those concerned to the comprehensive significance of the new possibilities for their lives. And on the topics of health and disease, life and death, and lastly leisure and work, symbols again came to be applied that corresponded with people's lives and which directly concerned their well-being and their dreams for the future.

The leaflet again brought all the objects in the exhibition together by showing how they contributed to the project.[25] On one page it was explained why the ants functioned as a model for human society and another half-page contained dialogues on the picture-statistical images.[26] Interestingly, neither a title nor a name for the exhibition can be found. Where we would expect it we find a pictorial chart (Figure 9) containing the key symbols of the exhibition. A figure resembling a cogwheel and divided in two depicts on the left, before a dark background, the bad past, illustrated by small houses with few and narrow windows and a child with rickets, while on the right we see, before a fair background, a bright future, illustrated by larger houses with broader windows and a healthy child. Below this image, just as in the sixth and twelfth chart, the direct form of address was used in order to connect with the visitors. Here the citizens were asked to make suggestions for the solution of the problem. The last sentence: "We want to know what you *think*.", addressed again and directly the observer or readers.

BILSTON – AN EXPERIMENT IN SOCIAL TECHNOLOGY?

The pamphlet used a pictorial chart for its motto, yet it contained no information on the question of how the exhibition was announced to the public or what it was called. Since Otto Neurath's interview with the *News Cronicle* in early December 1945 in Bilston, it was common knowledge that he objected to calling the project a social experiment. Instead he spoke of the "Bilston venture". The term experiment offended him, for "human beings do not like to be guinea pigs" (Baron 1945).

It was by no means unusual in 1940s and 50s Britain to employ the language of experiment. In the new "Mass Observation Studies" the colonization of new settlements (e.g. Stevenage) or the rate by which central institutions of health and leisure were frequented (e.g. Peckham Health Center London) were observed by mostly young social scientists and psychologist, many of them from the USA.[27] This context may explain why in the magazine *Discovery*, which was edited by Morley, there appeared under the title "An Experiment in Social Tech-

nology" a collective report about Bilston with contributions by Williams, Morley and Reilly (Williams 1947, Morley 1947, Reilly 1947), while at the same time Marie Neurath quoted pictorial statistics from the exhibition and the pamphlet in "The Journal of the Royal Institute of British Architects" under the title "An Isotype Exhibition on Housing" (Neurath 1947).

Had the two parties clashed on the issue of whether the Bilston project should be regarded as or packaged as a "socio-technological experiment"? The historical records do not settle this question conclusively. No doubt Marie Neurath asked Williams for a suitable suggestion for the motto of the project, which was also to appear on the guide to the exhibition.[28] Possibly, his evasive reply indicates that he was aware that in dealing with her, too, he could not use the word "experiment", although this was a concept associated with such schemes in the mind of the public. From the contemporary press reports one gets the impression that hardly anyone paid attention to the "Isotype Exhibition on Housing", for only the ants were named as the main attraction and figured prominently in the memory of the visitors. Yet this was not at all the way in which matters were portrayed in the subsequent reports and reminiscences of the organizers, which the literature on Bilston has taken for the sole truth about this case. Even Morley, who created this display, did not mention the ants afterwards. What had happened in the meantime? To not leave everyone guessing: the ants, which were meant to attract the visitors and to hold up a mirror before them became a threat to the prospects of the entire project and as a result needed to be deleted from the memory of all persons involved.

Marie Neurath's correspondence with Williams following the opening of the exhibition and their respective publications of 1947 elucidate this puzzling case of collective amnesia. Marie asked Williams repeatedly to tell her how the exhibition was received by the Bilstonians, for she herself had been able to attend only the opening.[29] Only after she had found out that the exhibition had been closed she heard back from Williams, who had stalled until at last reporting in March 1947 that the mood in the city council towards the project had suddenly changed. Out of the blue the "self-appointed mouth-piece of the other party rose to his feet and called the exhibition a 'flee circus'", declaring that no further funding would be allocated. He and the chairman of the municipal health committee had seen no alternative but to hold their peace and to close the exhibition.[30] In her reply Marie Neurath expressed disappointment, for she had hoped to pursue further Neurath's idea, which had been developed together with representatives of Bilston during his visit in 1945, to have the exhibition followed by a museum – as in Vienna – and to receive the support of Josef Frank, the architect of 'red' Vienna, for it.[31]

But the idea of using ants as an exhibit in a show about housing problems – an idea that was very debatable – was not the sole reason why the project failed. Williams also reported that the building of the new houses had to be stopped because financial support from London, on which the town had counted, did not arrive. After 165 houses had been completed the Ministry of Health withdrew its

support and its licence, because it no longer endorsed the "suction system of refuse collection and district heating".[32] Marie Neurath offered to use her connections with the Ministry, but Williams left Bilston soon after and took over the running of an altogether different district.[33]

Bilston shares its negative fame with other postwar utopias, which were buried as a result of the narrow financial margins within which Labour was forced to operate during its short term. In November 1946, when Williams composed his optimistic report on the Bilston project, he could not have predicted that half a year later the situation would be utterly hopeless.[34] Marie Neurath, who wrote her report on the pictorial statistics in May 1947, lost no words on the circumstances of the project.[35] With hindsight and knowing how swiftly the project failed, the report of 1947 reads like a collective dream about the bright, healthy and happy future of Bilston. This illustion the dreaming utopians, Marie Neurath, Williams, Reilly and Morley, were reluctant to let go.

NEURATH'S ISOTYPE IN THE BRITISH CONTEXT OF PLANNING, DEMOCRATIZATION AND POPULARIZATION

The case of Bilston was neither unusual nor a rare exception. The desires, hopes and expectations formulated here typified the vision of postwar reconstruction that evolved around 1945 in Great Britain.[36] Anticipating Allied victory, these plans made the active collaboration of citizens a condition for the preservation of social peace even after the war had ended.[37] Labour pleaded for public investments in the sector of health and housing. It was also not unusual to express the experience of the past and the goals of the future in the language of pictorial statistics. Prominent among the many examples are classical pieces of promotional literature like Beveridge's social plan, published by Sir Donald Davidson under the title *Social Security* (Davidson 1943), like John Boyd Orr's plan for world nutrition, *Food and the People* (Boyd Orr 1943), or Stephen Taylor's *The Battle for Health* (Taylor 1944). Worth mentioning here are also *Interglossa*, the proposal of the biomathematician Lancelot Hogben to develop a new planned language (Hogben 1944), and Paul Rotha's films on nutrition and housing, *The World of Plenty* and *The Land of Promise*, which revolutionized British documentary film.[38] Neurath's graphical technology for representing statistics was particularly attractive to Liberals and to Left activists, who were concerned to equip citizens with empirical facts about economics, society and health. Nevertheless one wants to know how Neurath the émigré interacted with these attempts (so similar to his own) to improve communication between different strata of society, and in what ways he inspired his British colleagues. There is no time here to discuss Hogben and Rotha in great detail. Both were members of the advisory committee of the ISOTYPE Institute and they were close to Neurath, yet the scholarly evaluation of their relationship with Neurath has not yet begun.[39]

When he embarked on linguistic theory, Hogben already was the well-known author of best-selling popular works, notably *Mathematics for the Million* (1936) and *Science for the Citizen* (1938). The secret of his success as a popularizer was to keep abstract knowledge brief and to make more room for applied knowledge instead. The audience of *Interglossa*, however, were the popularizers themselves. Hogben tackled the old question of what kind of language would be best suited as a global language, discussing first the attempt of the linguist C.K. Ogden to reduce English to 850 words, "Basic English".[40] To use English as the basis for a world language was undemocratic, Hogben thought: "A satisfactory auxiliary must be everybody's language because it is also nobody's language" (Hogben 1944, 9). His *Interglossa* was a more logical version of Esperanto, he argued, which could, moreover, be acquired with the aid of visual languages like ISOTYPE, a claim he proceeded to substantiate in his book.

Hogben's criticism of Ogden complemented Neurath's ideas, although Neurath had in fact written two handbooks in Basic English on the use of ISOTYPE (Neurath 1936, 1937). Yet in contrast to Hogben Neurath saw in *Interglossa* also an example of how adult education was still based on the assumption that in the future world order everyone already could read or write: "Colonial people who do not understand one of the languages which can transfer scientific knowledge like English or Chinese, Basic or Interglossa, are badly off in the educational field" (Neurath [ca.1945] 1996, 256). Neurath worked from the lowest level of education, hoping in this manner to defuse frustration and feelings of inferiority on the part of those concerned. For this reason he took the step towards visual communication and developed ISOTYPE as "a kind of visual lingua franca" (ibid., 264).

In his incomplete manuscript: *Visual Education. Humanisation versus Popularisation*, Neurath rejected the idea of "popularisation" and promoted the idea that knowledge and information required "humanizing". In her report on Bilston, Marie Neurath made the common point that the purpose of their exhibition had not been to spread the illusion that a layperson could step into the shoes of the expert following their instruction: "We did not want to give a popularized version of what a student of medicine has to learn; we did not want to create the illusion as if, after having seen the exhibition, the visitors would be able to find out about their health and cure themselves" (Neurath 1947, 600). Like many other popularizers of their era the Neuraths were concerned not to produce yet more specialists but to supply people with information that would help them take charge of their lives. Their interests concerned generally valid knowledge that was relevant for everyday life. In the quest for symbols, objects were chosen with which people were familiar from their everyday lives, like (as we saw on the example of Bilston) house, bed, garden, sun, factory, football pitch, and so on. The condition for humanizing knowledge was a language that was precisely not a specialist language of the few. Moreover, to humanize statistics was coextensive with making disappear the results of social research and therefore the numbers behind the symbols, by reducing uncountable or incomprehensible sta-

tistics to the dimensions in which counting was performed by those who usually used the ten fingers of their hands to do so.

ISOTYPE was meant to facilitate dialogue between teachers and pupils, between experts and laypersons. This can be seen for example in the role visual statistic played in the documentary films of Paul Rotha. In *The World of Plenty*, for example, the voice of the nutritionist Boyd Orr did not figure as that of an expert who can make the figures behind nutritional statistics talk. There also was the role of the "man-in-the-street", who liked this system and who wanted to have everything explained in it. So when Boyd Orr explains the complex cycle of food production and its distribution, the man-in-the-street suggests that Boyd Orr should rather use ISOTYPE so he can understand what is being explained (Knight and Rotha 1945). ISOTYPE entered not as a new expert language or a secret code, but as a means to motivate people to appropriate knowledge and information themselves. We should note though that Eric Knight, who wrote the text for Rotha's films, appeared to advocate the concept of popularization that had been rejected by Neurath. Knight made the "First Announcer" comment on Boyd Orr's views on a juster distribution of food as follows: "Science has the answer! We know now what can be done!" (ibid., 47). Presumably Neurath's rendition of scientific arguments through ISOTYPE would have turned out more modest, for in his view ISOTYPE did not make such arguments more compelling but only more comprehensible.

Neurath appears to have been who Rotha had been looking for all those years. Immediately upon his arrival in Oxford Rotha showed up and offered Neurath to collaborate with him on films (Neurath and Cohen 1973, 72). That Rotha absorbed Neurath's ideas on visual pedagogy like a sponge transpires in his lecture on "Films and other Visual Techniques in Education", given shortly after Neurath's death before the Museums Association. On this occasion he canvassed for Neurath's ISOTYPE on the grounds that they helped him realize the "clear-cut visual communication" which became his hallmark as a film maker (Rotha [1946] 1953, 89). Using almost Neurath's exact words, he added symbols and animated pictorial statistic as the indispensable element of "our orchestra for visual techniques" (ibid., 91), thus giving them a place in a tradition. Later, Hogben likewise saw things in this way. In his book, *From Cave Painting to Comic Strip*, a world history of visual communication that was published by the ISOTYPE Institute in 1949, we find Neurath classified among the contributions of modernity in a chapter entitled "Standardisation, Stereotype and Isotype" (Hogben 1949, 183).

CONCLUSION

The exhibition on the planning of Bilston was not the first of its kind in Britain. Films about housing problems, new housing and health were not a novelty.[41] Since the 1940s plans for reconstruction were publicly discussed and displayed. However, in exhibitions like "Rebuilding Britain", which was held in 1943 in the National Gallery, the displays had consisted of architectural drawings and models both of which probably remained incomprehensible to the lay public. In the films, moreover, the planners paraded themselves as infallible scientific experts, just as had been the practice of the master planner Patrick Abercrombie.[42]

With the Neurath's arrival ISOTYPE became an instrument for communicating British "Postwar Utopianism" in the area of planning, democratization and popularization in a special way. Rotha's film about housing problems, *The Land of Promise* (1945), not only conveyed facts through animated picture statistical diagrams, but also, ISOTYPE received a life if its own. The method became personified in "Mister Isotype", who was inserted into the script. Like Neurath Rotha was not concerned to let images speak for themselves, a rhetoric frequently used to justify visualization. Rather, "Mister Isotype" was used to explain how ISOTYPE worked, for observers first needed to get to know the visual vocabulary and how it was composed before they could understand what the planners or Neurath wanted to show them.

From recent research we have learnt that Neurath first attempted to use the picture statistical method internationally at the international congresses of modern architects in the 1930s. These architects looked to Neurath for a universal language that translated social planning into building plans for the towns of the future. But their collaboration with Neurath failed: he disappointed the architects, who expected from him not only a universal language for communication, but universal scientific criteria legitimate their plans (Chapel 1996, Faludi 1996). In Bilston, too, Neurath was in some sense confronted with the expectation that he would give his blessing to final and rational plans, over the heads of the citizens. The ants, if we recall, were meant to show people the model which, from the standpoint of natural history, demonstrated the most sensible model for living together. Neurath would not have approved of such dictatorial planning. Planners as well as scientists should not as a rule be given the last word, he believed, for they, too, were able to take only one of many possible and rationally justifiable standpoints. For Neurath democracy had "a certain tendency to orchestrate different opinions and different ways of life" (Neurath [ca. 1945] 1996, 249). Unlike others he did not speak of "planning" in general and of "social planning"; rather, he always used the expression "democratic planning" so as to remind planners that those concerned must be included into the planning process for it to be successful.[43] His own contribution was the creation of a standardized basis for communication between different nations, different levels of education and dif-

ferent fields of expertise. In this way ISOTYPE could contribute to the democratization of expert knowledge, because his system tackled the communication problem between experts and lay people. His pictorial statistics served as a means to balance arguments on the same level. The question of whether life in a community was the right thing for everyone was to be left to the reasoning of neither architects nor psychologists nor economists to decide. Neurath pictured the planning of democracy to function like an orchestra, where different instruments played together according to their potential.

Not all questions in a democracy, he was convinced, required a majority decision. And after hearing of the case of Bilston we should not be surprised to find an example of town planning in his manuscript: "If city planning is being discussed, and one third of the population want small houses with gardens, while two thirds want big flats, why should the society decide by a majority vote about either?" (Neurath [ca. 1945] 1996, 249). All attempts at enlightenment succeeded when they gave a voice to people in which they were able to formulate their needs and which were then also understood by others, for example, experts. As an example of how democracy can function in practice he referred to the tried and tested English metaphor of "muddling through": "One can hardly imagine that rigid principles will prevail; rather a gigantic 'muddling through' may be the characteristic of a future world community in which orchestration will play a greater part than unification" (ibid., 334).[44]

In ISOTYPE, Neurath saw the programme of the Enlightenment completed in the "Century of the Eye", as he repeatedly called the 20th century. In contrast to his 18th-century predecessors he did not talk down from the pulpit. His was not the role of the theologian or teacher, for he was not a moralist.[45] Neurath did not lay down what to think, but with what tools thinking should take place. His role was that of a conductor who proposed a rhythm and who made sure that the nations of the new world order will play – each for itself and yet all of them together.

Translation by Anna-K. Mayer.

ACKNOWLEDGEMENTS

My research on Bilston was conducted with the aid of a History of Medicine Wellcome Research Fellowship in 1999-2000 at the Department of History and Philosophy of Science, University of Cambridge, UK. I want to take this opportunity to thank the Wellcome Trust and my sponsor, Nick Hopwood. For access to, help in evaluating archival material on Bilston in and for permission to reproduce the charts from the collections of the ISOTYPE Institute at the Department of Typography and Graphic Communication, University of Reading, UK, I am

indebted to Chris Burke. Preliminary results of this study were presented at the conference on "Pursuing the Unity of Science. Ideology and Scientific Practice in the Interwar Period" at the University of Maastricht, The Netherlands, in July 2000, and in November 2002 in the workshop on "Otto Neurath im Kontext" of the Graduiertenkolleg "Auf dem Weg in die Wissenschaftsgesellschaft" at the Institut für Wissenschafts- und Technikforschung, University of Bielefeld. I thank Lars Bluma, Christine Brecht, Martin Carrier, Andreas Faludi, Harmke Kamminga, Anna-Katharina Mayer, Nader Vossoughian, Diederick Raven, Friedrich Stadler and Michael Stöltzner for helpful comments.

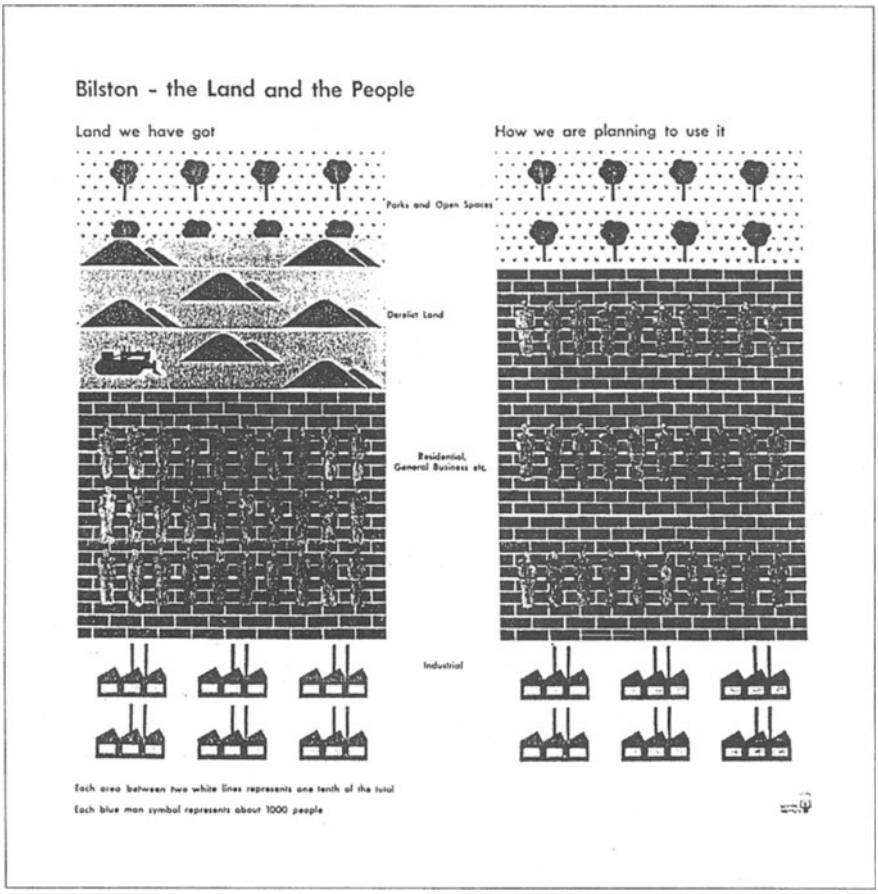


Figure 2

Chart 2, IC (MS 1091 3.2/93-96)

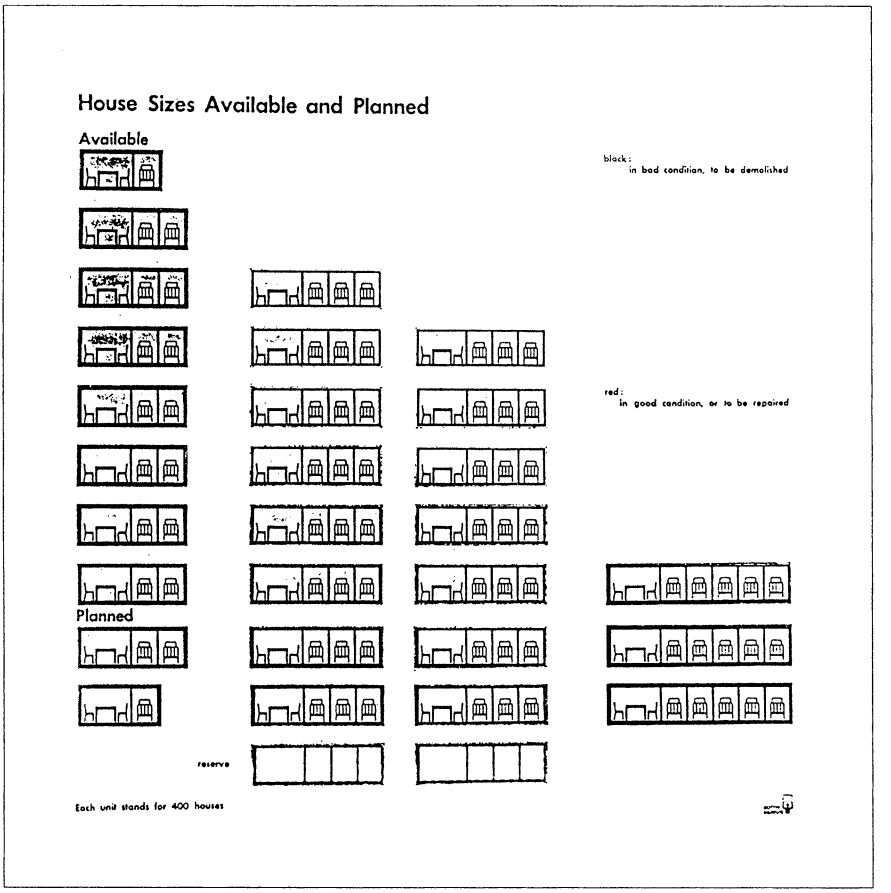


Figure 3

Chart 4, IC (MS 1091 3.2/93-96)

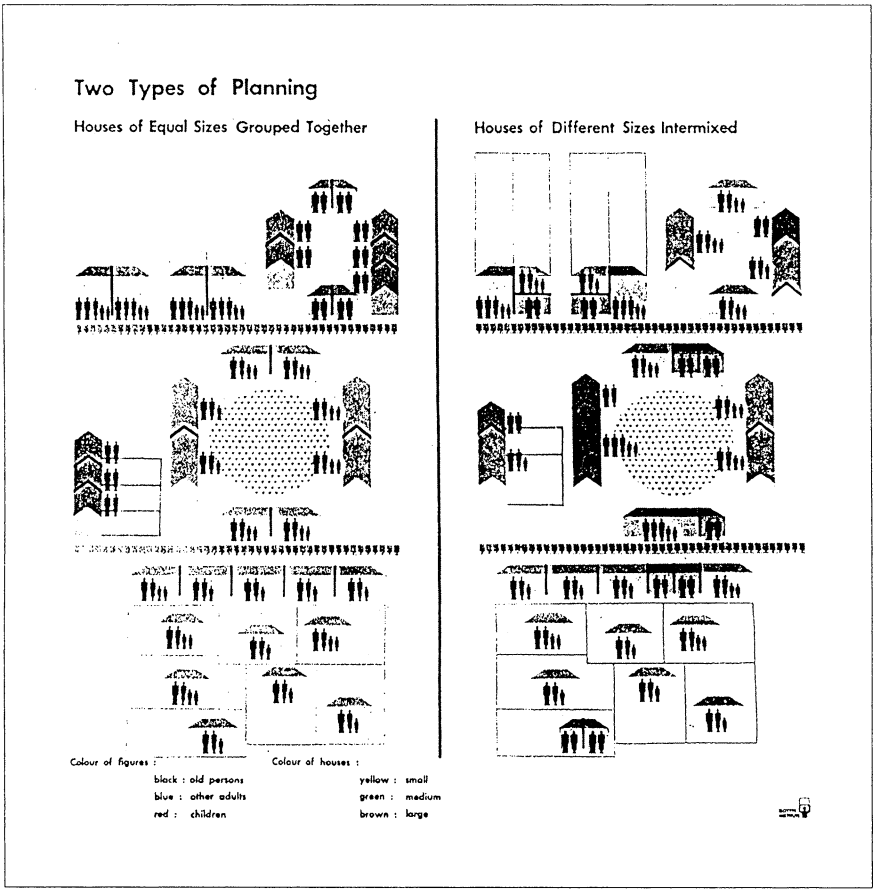


Figure 4

Chart 5, IC (MS 1091 3.2/93-96)

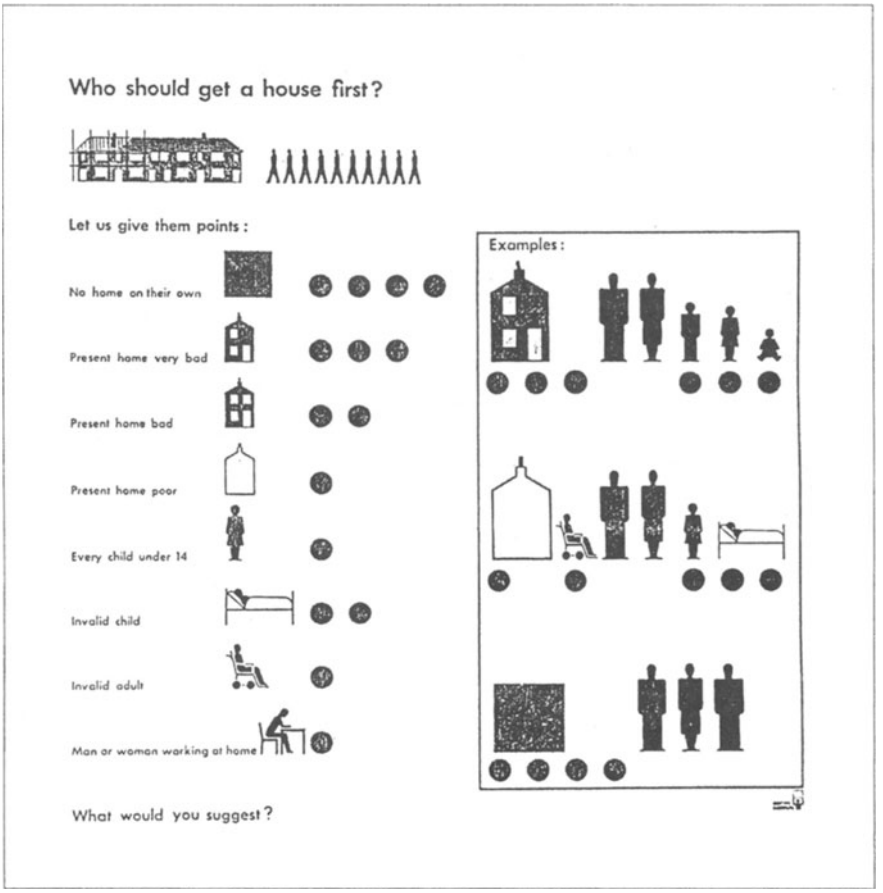


Figure 5

Chart 6, IC (MS 1091 3.2/93-96)

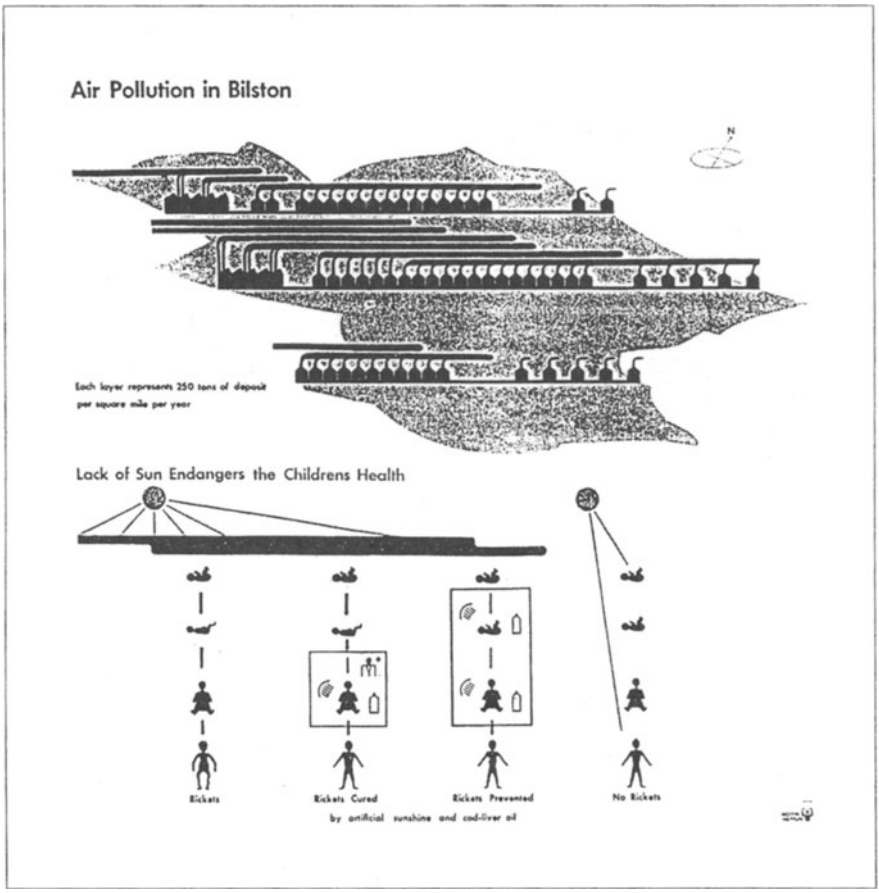


Figure 6

Chart 7, IC (MS 1091 3.2/93-96)

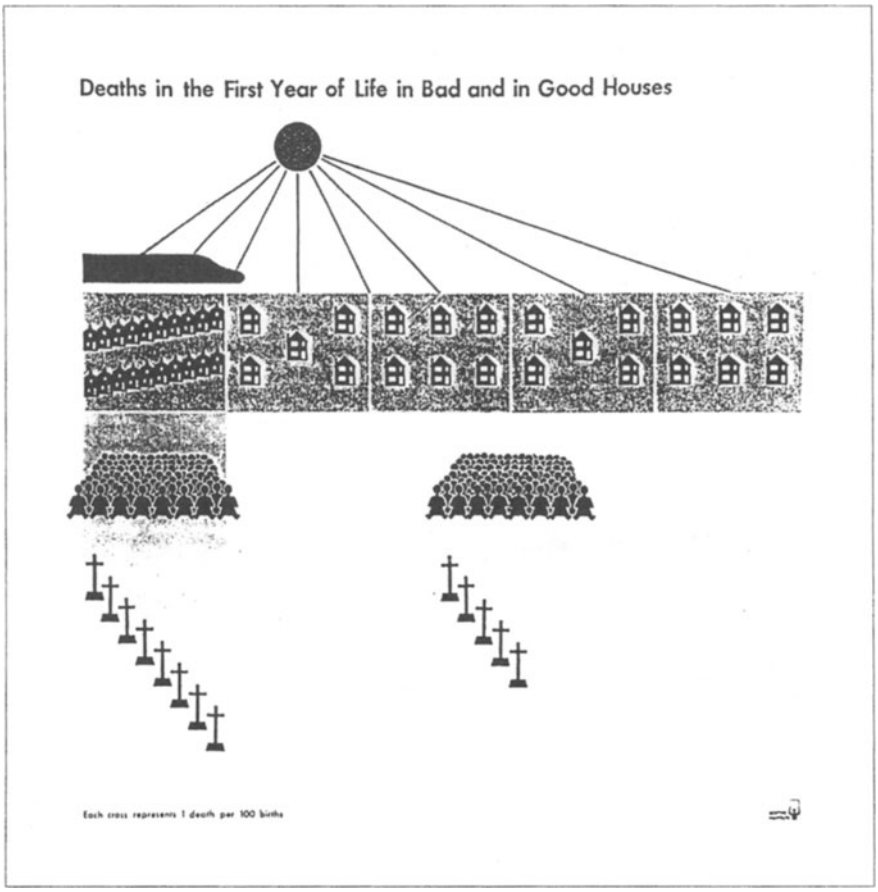


Figure 7

Chart 9, IC (MS 1091 3.2/93-96)

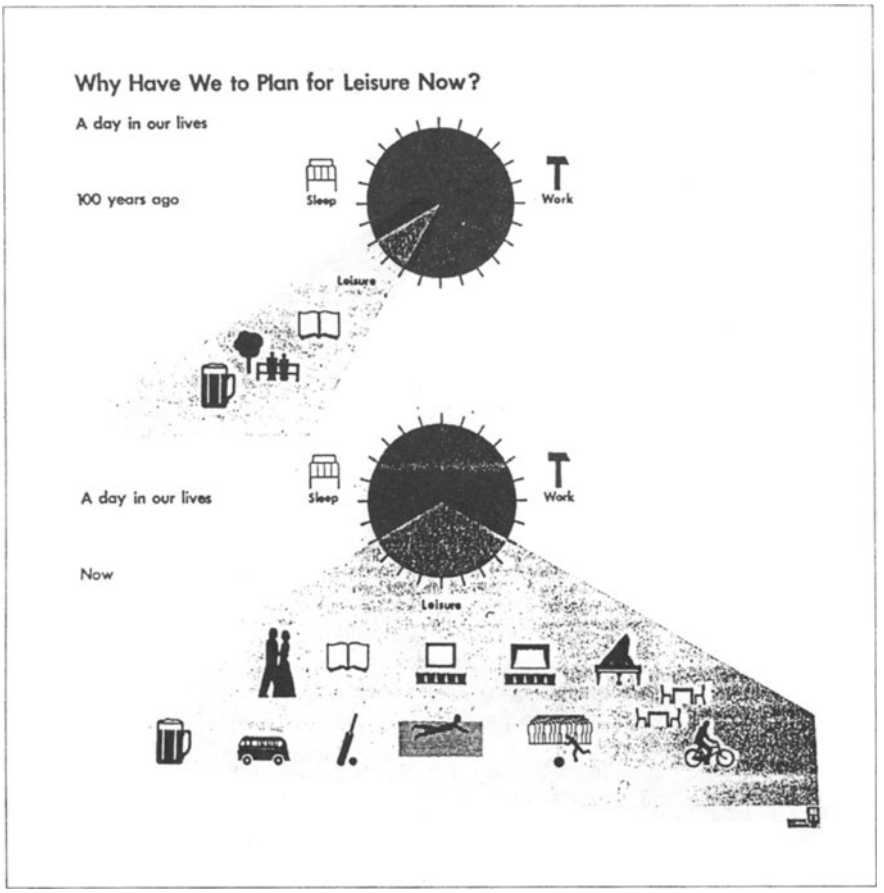


Figure 8

Chart 12, IC (MS 1091 3.2/93-96)



BILSTON has many problems, you know what they are. Perhaps you have some suggestions for solving them. This exhibition shows a few of the problems and suggests some answers. If you will help by visiting it and thinking about what you see, we can all work together to build your Bilston.

READ this and discuss it with your friends. Think our your suggestions and bring them along.

WE want to know what you *think*.

Figure 9

Front Cover of the Pamphlet
IC (MS 1091 3.2/93-96)

NOTES

1. Marie Neurath recalls this in: Neurath and Cohen 1973, 63-64. For a survey over the various exhibitons and publications see Stadler 1982. For the collaboration with Arntz, see also Gerd Arntz 1976, Politische Konstruktivisten 1975.
2. I use the most comprehensive version of this manuscript, discovered and edited by Juha Manninen: Neurath [ca. 1945] 1996.
3. Isotype Collection, Department of Typography and Graphical Communication, University of Reading, henceforth IC.
4. For the period of the industrialization, see also Price 1951, 14-29.
5. On this context see also Ward 1994, 80-115.
6. On Reilly's key achievement see also Charles Reilly 1997, on the "Really Green" ibid, 162 and Matless 1998, 237, on Community Architecture: Reilly 1944, on Bilston: Reilly 1946. For an earlier example of using community-arguments in the planning of urban housing, see Darling 2000.
7. "The Battle of Birkenhead", *Architects' Journal*, 12 October 1944, 264 and "Birkenhead. Community versus Segregation", *Architects' Journal*, 3 August 1944, 85-87, for subsequent alterations of the plan "Birkenhead, a new plan by Reilly and Aslan", *Architects' Journal*, 9 May 1946, 359-362, and for letters from readers see *Architects' Journal*, 12 September 1946 and 10 October 1946, and *Picture Post*, 22 July 1944.
8. See e.g. the debate on Abercrombie's plan for Greater London, in Gold 1997, 179 and Ward 1994, 97-99.
9. "How to get the Houses. The Outline of a Plan", *Picture Post*, 14 July 1945, 16-17, 26. "The Plan Matures", *Picture Post*, 25 August 1945.
10. See "Civic Survey Exhibition Sept. 1944 in Town Hall", pamphlet, IC (MS 1091 3.2/93-96).
11. Radio conversation with Dr Robert Abbott on the BBC Midland Home Service, Thursday, 7 November 1946, 6.20-6.25 pm, typewritten manuscript, IC (MS 1091 3.2/93-96)
12. Neurath ([1926a] 1991) was responsible for the exhibition of the Austrian Social Insurance in the Austrian Pavillion at the GESOLEI, a large exhibition on health care, social welfare and physical exercise, which took place in Düsseldorf in 1926 and of which it was said that it attracted over 7 million visitors. On visual statistics as a means to persuade and communicate health education in exhibitions, see Brecht/Nikolow 2000, Nikolow 2001, 2002a, 2002b.
13. Williams to Neurath, 4 October 1946, IC (MS 1091 1/14-17). Subsequently, Morley became a wellknown entomologist and sociobiologist. For him a formicarium became an instrument for observing the social behaviour of ants. See Morley 1953 with a preface by e.d. Julian Huxley.
14. The caricatures were by Dodie Mastermann; see Neurath to Williams, 26 October 1946, IC (MS 1091 1/14-17), which reports on the meeting with Morley and Mastermann and on the work on the pamphlet. See also the undated manuscript: "The Bilston Venture", signed by Morley, IC (MS 1091 1/14-17).
15. Neurath to Williams, 5 October 1946, IC (MS 1091 1/14-17). She recommended photos as in Abrams 1945.
16. Neurath to Morley, 12 November 1946, IC (MS 1091 1/14-17).
17. Undated pamphlet, IC (Box "The Bilston Experiment" MS 1091 3.2/93-96).
18. On the term: "transformation" see Neurath [1933] 1991, 239. On Marie Reidemeister's employment in Vienna see her reminiscences in Neurath and Cohen 1973, 56-58.
19. These can be found in IC, also the location of the figures reproduced here.
20. Neurath to Williams, 30 July 1946, IC (MS 1091 1/14-17)
21. The question here was: how can there be a universal symbol for mortality if one wants to speak in the name of all cultures? In pictures for and about the USA Neurath frequently used urns. See Neurath [ca. 1938] 1991, 428, 429, 444 and Neurath (1944), 64. Although Neurath aimed to standardize his symbols in practice he departed from this principle. As in other representation from this period he used crucifixes with tombstones for deaths of tuberculosis, heart disease (siehe Taylor 1944, table IV, 68-69) and diphtheria (ibid. table V, 73, chart VI, 77), but cruci-

fixes without tombstones for infant mortality (see ibid. table XII, 101) and for all causes of death and all age groups (see ibid. table II, 24-25). Marie Neurath may have referred to earlier representations, e.g. Neurath [1939], 1991, 491. For a brief period during his career Neurath used coffins to communicate mortality, then he announced his intention to use it only for infant mortality in future, yet this did not last. See above, and for his announcement see Neurath [1926b], 1991, 51, 51-55.

22. He explained this as a principle of design (Neurath [ca. 1938] 1991, 436). For examples see ibid., 426, 435, 438, 442, 448.
23. He dedicated an entire chapter to the everyday life of man ("Alltagsleben der Menschen"), see Neurath [1939] 1991, 524-563.
24. Ibid., 546.
25. A copy of the pamphlet can be found in IC. Marie Neurath quoted from it in her article on Bilston (Neurath 1947).
26. For the dialogues see exerpts in Neurath 1947.
27. On Stevage, see Orlands 1952; on Peckham see Pearse and Crocker 1943; also Beach 2000.
28. Neurath to Williams, 26 October 1946, Williams to Neurath, 31 October 1946, both IC (MS 1091 1/14-17).
29. Neurath to Williams, 6 January 1947, 5 March 1947, 19 March 1947, all IC (MS 1091 1/14-17).
30. Williams to Neurath, 27 March 1947, IC (MS 1091 1/14-17).
31. Neurath to Williams, 18 March 1946, 6 January 1947, 19 March 1947, 8 May 1947, all IC (MS 1091 1/14-17). Joseph Frank, the brother of the physicist Phillip Frank (of the Vienna Circle), lived then in exile in New York. Back in Vienna, Neurath had obtained Frank's services for the design of his exhibitions and while in exile the two entertained a lively correspondence. For the original plan see also Baron 1945.
32. Williams to Neurath, 27 March 1947, IC (MS 1091 1/14-17), for the technical details of the plan see Reilly 1947.
33. Williams to Neurath, 30 April 1947, Neurath to Williams, 8 May 1947, 21 Mai 1947, 27 August 1947, IC (MS 1091 1/14-17).
34. Neurath to Williams, 25 November 1946, IC (MS 1091 1/14-17).
35. Williams to Neurath, 21 May 1947, IC (MS 1091 1/14-17).
36. The story of these unsuccessful projects has – in spite of their large number – not yet been written. On the failure of the Labour Party in the question of housing see Ravetz 1995 and on the rejections of Rotha's documentary film, *The Land of Promise*, on the part of the Ministry of Health in 1945, on the grounds of his radical demand for 4 million new houses to be built, see Haggith 1990.
37. For a brief discussion of surprisingly similar themes in the context of postwar reconstruction in the aftermath of the Great War, see Lawrence and Mayer 2000.
38. Although there is Tim Boon's study of *The World of Plenty* (Boon 1997), we are still waiting for an answer to the question of why Rotha and others looked to ISOTYPE in particular as the suitable means to communicate their ideas to the public and thus to decide against other methods.
39. For biographical details on Hogben see Werskey [1978] 1988, Hogben 1998 and for Rotha see Rotha 1973. The collaboration with the ISOTYPE Institute does not form part of these recollections which were written during the Cold War. There are numerous references to it in the correspondence of these individuals with Neurath at the time, which, however, has not yet been studied.
40. In 1943 the British government seriously considered to make Basic English the second language for those whose native tongue was not English, particularly in the colonies. On this see Ogden 1943 and for public response in letters to the *Picture Post*, see Ibid., 19 November 1943, 3.
41. For a survey on exhibitions see Gold 1997, 116-140 and on documentary films see Haggith 1990, Gold and Ward 1997.
42. For "Rebuilding Britain", see *The Journal of the Royal Institute of British Architects*, February 1943, 75-84, 97-102 and the special issue of *The Architectural Review*, April 1943, 85-112; for the planners on the screen see Gold and Ward 1997, 73-76; for Abercrombie see Dix 1981.
43. Julian Huxley, another member of the advisory committee of the Isotype Institute, also used the term "democratic planning" in order to stress that planning for those concerned was possible

only with their collaboration. On this see his report on the Tennesse Valley Authority Project in the USA (Huxley 1943).

44. On Neurath's planning philosophy see also Faludi 1996 who interprets it as pluralistic.
45. He referred to Basedow's *Elementarwerk* which dates from the end of the 18th century, one of the earliest handbooks for teachers which is complemented by a collection of pictures. Neurath criticised its selection for its utter failure to adapt to the needs of children. For picture encyclopedias and their role in German enlightenment pedagogy, see te Heesen [1997] 2002.

REFERENCES

M. Abrams, "Housing and Homes in Post-war Britain", in: *Future Books*, I, 1945, pp. 41-49.

Anonymous, "Ants will guide this town to healthier living", in: *Birmingham Gazette*, 2 November 1946.

St. Baron, "Man with a lot of Happiness", in: *News Chronicle*, 4 December 1945, p. 2.

Abigail Beach, "Potential for Participation: Health Centres and the Idea of Citizenship, c. 1920-1940", in: Christopher Lawrence and Anna-Katharina Mayer (Eds), *Regenerating England: Science, Medicine and Culture in Interwar Britain*. Amsterdam: Rodopi 2000, pp. 203-230.

Bilston, Tettenhall and Wednesfield (1998), Compiled by Mary Mills and Tracey Williams. Stroud 1998.

Tim Boon, "Agreement and Disagreement in the Making of World of Plenty", in: David F. Smith (Ed), *Nutrition in Britain. Science, Scientists and Politics in the Twentieth Century*. London and New York: Routledge 1997, pp. 166-189.

Sir John Boyd Orr, *Food and the People*, in: Ch. Madge (Ed), *Target for Tomorrow* 3, London: The Pilot Press 1943.

Christine Brecht and Sybilla Nikolow, "Displaying the Invisible. 'Volkskrankheiten' on Exhibition in Imperial Germany", in: *Studies in the History and Philosophy of Biological and Biomedical Sciences* 31, 2000, pp. 511-530.

Charles Reilly & The Liverpool School of Architecture 1904-1933, Catalogue of an exhibition at the Walker Art Gallery, Liverpool 2 October 1996 - 2 Februar 1997, Liverpool: Liverpool University Press 1997.

Eric Chapel, "Otto Neurath and the CIAM – The International Pictorial Language as a Notational System for Town Planning", in: Elisabeth Nemeth and Friedrich Stadler (Eds), *Encyclopedia and Utopia*, Dordrecht, Boston, London: Kluwer 1996, pp. 167-182.

Elizabeth Darling, "'Enriching and Enlarging the Whole Sphere of Human Activities': The Work of the Voluntary Sector in Housing Reform in Inter-War Britain", in: Christopher Lawrence and Anna-Katharina Mayer (Eds), *Regenerating England: Science, Medicine and Culture in Interwar Britain*, Amsterdam: Rodopi 2000, pp. 149-178.

Sir Ronald Davidson, *Social Security. The Story of British Social Progress and the Beveridge Plan*, London, Toronto, Bombay, Sydney: George G. Harrap and Company 1943.

G. Dix, "Patrick Abercrombie, 1879-1957", in: Gordon E.Cherry (Ed), *Pioneers in British Planning*, London: Architectural Press 1981, pp. 103-130.

Maurice Edelman, "Planning Post-War Britain. The Example of Birkenhead", in: *Picture Post*, 24, 2, 8 July 1944, pp. 16-20.

Andreas Faludi, "Otto Neurath and Planning Theory", in: Elisabeth Nemeth and Friedrich Stadler (Eds), *Encyclopedia and Utopia*, Dordrecht, Boston, London: Kluwer 1996, pp. 201-213.

Gerd Arntz. Kritische Grafiek en Beeldstatistiek, The Hague: Haags Gemeentemuseum 1976.

John R. Gold, *The Experience of Modernism. Modern Architects and the Future City, 1928-53*, London, Weinheim, New York: E & FN Spon 1997.

John R. Gold and Stephen V. Ward, "Of Plans and Planners. Documentary Films and the Challenge of the Urban Future, 1935-52", in: David B. Clarke (Ed), *The Cinematic City*, London: Routledge 1997, pp. 59-82.

Graphic Communication through Isotype, Reading: University of Reading 1975.

T. Haggith, "Post-war reconstruction as depicted in official British films of the Second World War", in: *Imperial War Museum Review* 7 ,1990, pp. 34-55.

Anke te Heesen, *The World in a Box. The Story of an Eighteenth-Century Picture Encyclopedia*, Chicago, London: University of Chicago Press 2002, (Translated from German *Der Weltkasten. Die Geschichte einer Bildenzyclopädie aus dem 18. Jahrhundert*, Göttingen: Wallstein 1997).

Adrian and Anne Hogben, *Lancelot Hogben. Scientific Humanist. An unauthorised Autobiography*, Rendlesham: Merlin Press 1998.

Lancelot Hogben, *Interglossa. A Draft of an auxiliary for a democratic world order, beeing an attempt to apply semantic principles to language design*, London: Pelican Books 1944.

Lancelot Hogben, *From Cave Painting to Comic Strip. A Kaleidoscope of Human Communcation*, London: Max Parrish & Co. 1949.

Julian Huxley, "Democratic Planning", in: *The Architectural Review*, 93, April 1943, pp. 139-166.

Eric Knight and Paul Rotha, *World of Plenty. The Book of the Film*, London: Nicholson & Watson 1945.

Hadwig Kräutler, "'Scientific World Conception' and Museum Planning", in: Elisabeth Nemeth and Friedrich Stadler (Eds), *Encyclopedia and Utopia*, Dordrecht, Boston, London: Kluwer 1996, pp. 191-200.

Christopher Lawrence and Anna-Katharina Mayer, "Regenerating England: An Introduction", in: idem (Eds.), *Regenerating England: Science, Medicine and Culture in Interwar Britain*, Amsterdam: Rodopi 2000, pp. 1-23.

David Matless, *Landscape and Englishness*, London: Reaktion Books 1998.

Derek W. Morley, "Sociological Approach to Town Planning: The Bilston Experiment – A Problem in Social Engineering", in: *Discovery*, August 1947, pp. 250-251.

Derek W. Morley, *Ants*, London: Collins 1953.

Otto Neurath, "Österreichs Sozialversicherung auf der 'Gesolei' in Düsseldorf", in: *Österreichische Gemeinde-Zeitung*, 3, No. 10, 1926a, pp. 312-317 (Reprinted in Neurath 1991, 28-33).

Otto Neurath, "Schwarzweißgraphik", in: *Österreichische Gemeinde-Zeitung*, 3, No. 10, 1926b, pp. 334-338 (Reprinted in Neurath 1991, 51-55).

Otto Neurath, O, "Soziale Aufklärung nach Wiener Methode", in: *Mitteilungen der Gemeinde Wien*, No. 100, 1933, pp. 25-33 (Reprinted in Neurath 1991, 231-239).

Otto Neurath, *International Picture Language. The First Rules of Isotype*, London: Kegan Paul 1936. (Reprinted and translated as *Internationale Bildsprache* in Neurath 1991, 355-398).

Otto Neurath, *Basic by Isotype*, London: Keagan Paul 1937.

Otto Neurath, "Health Education by Isotype" together with Herbert E. Kleinschmidt, in: *American Public Health Association*, ca. 1938, (Reprinted and translated as "Gesundheitserziehung mit Isotype" in Neurath 1991, 421-448).

Otto Neurath, *Modern Man in the Making*, New York: Albert Knopf 1939 and London: Secker and Warburg 1939, (Reprinted and translated as *Auf dem Wege zum modernen Menschen* in Neurath 1991, 449-590).

Otto Neurath, "The Human Approach to Visual Education", in: *Health Education Journal*, 2, 2, 1944, pp. 61-66, (Reprinted and translated as "Der humane Zugang zur Bildpädagogik" in Neurath 1991, 618-829).

Otto Neurath, *Gesammelte bildpädagogische Schriften*. Rudolf Haller and Robin Kinross (Eds), Wien: Hölder-Pichler-Temsky 1991.

Otto Neurath, "Visual Education. Humanisation versus Popularisation", in: Elizabeth Nemeth and Friedrich Stadler (Eds), *Encyclopedia and Utopia*, Dordrecht, Boston, London: Kluwer 1996, pp. 245-335.

Marie Neurath, "An Isotype Exhibition on Housing", in: *The Journal of the Royal Institute of British Architects*, October 1947, pp. 600-603.

Marie Neurath and Robert S. Cohen (Eds), *Otto Neurath. Empiricism and Sociology*, Dordrecht-Holland, Boston-USA: D. Reidel Publishing Company 1973.

Sybilla Nikolow, "Der statistische Blick auf Krankheit und Gesundheit. 'Kurvenlandschaften' in Gesundheitsausstellungen am Beginn des 20. Jahrhunderts in Deutschland", in: Ute Gerhardt, Jürgen Link and Ernst Schulte-Holtey (Eds), *Infographiken, Medien und Normalisierung. Zur Kartografie politisch-sozialer Landschaften*, Heidelberg: Synchron 2001, pp. 223-241.

Sybilla Nikolow, "Die graphisch-statistische Herstellung der Bevölkerung. Bevölkerungskonzepte in der Gesundheitsaufklärung in Deutschland vor 1933", in: Rainer Mackensen (Ed), *Bevölke-*

rungslehre und Bevölkerungspolitik vor 1933, Opladen: Leske + Budrich 2002a, pp. 297-324, 283, 286.

Sybilla Nikolow, "Anormale Kollektive. Die Darstellung des 'Altersaufbaus der Bevölkerung des Deutschen Reiches' auf der Gesolei von 1926", in: Herbert Körner and Angela Stercken (Eds), *1926-2002. Ge So Lei. Kunst, Sport und Körper*, Ostfildern-Ruit: Hatje Cantz 2002b, pp. 217-221, 225-226.

Charles K. Ogden, "Can Basic English be a World Language? ", in: *Picture Post*, 23 October 1943, pp. 23-25.

Harold Orlands, *Stevenage. A Sociological Study of a New Town*, London: Routledge & Kegan Paul 1952.

Innes H. Pearse and Lucy H. Crocker, *The Peckham Experiment. A Study in The Living Structure of Society*, London: George Allen and Unwin 1943.

Politische Konstruktivisten. Die "Gruppe progressiver Künstler" Köln, Berlin: Neue Gesellschaft für bildende Kunst 1975.

John M. Price, *The Story of Bilston*, Bilston: John Price & Sons 1951.

Alison Ravetz, "Housing the People", in: John Fyrth (Ed), *Labour's Promised Land? Future and Society in Labour Britain 1945-1951*, London: Lawrence & Wishart 1995, pp. 146-162.

Sir Charles H. Reilly, *Architecture as a Communal Art*, London: B.T. Batsford 1944.

Sir Charles H. Reilly, "Bilston Communities", in: *New Statesman & Nation*, 23 November 1946, p. 380.

Sir Charles H. Reilly, "A Town Planner's Solution", in: *Discovery*, August 1947, p. 253.

Paul Rotha, "Films and Other Visual Techniques in Education, 1946", in: *Rotha on the Film. A Selection of Writings about the Cinema*, London: Faber and Faber 1953, pp. 88-97.

Paul Rotha, *Documentary Diary*, London: Seeker & Warburg 1973.

Friedrich Stadler, *Arbeiterbildung in der Zwischenkriegszeit. Otto Neurath-Gerd Arntz*, Wien, München: Löcker Verlag 1982.

Stephen Taylor, *Battle for Health. A primer of Social Medicine*, London: Nicholson & Watson 1944.

Stephen V. Ward, *Planning and Urban Change*, London: Paul Chapman Publishing Ltd. 1994.

Gary Werskey, *The Visible College. A Collective Biography of British Scientists and Socialists of the 1930s*, 2nd ed. 1988, London: Free Association Books 1978.

A.V. Williams, "Talk on Public Education in Planning", in: *The Architects' Journal*, 1 November 1945, p. 328.

A.V. Williams, "The Details of the Problem", in: *Discovery* August 1947, pp. 251-253.

Lawrence Wolfe, *The Reilly Plan. A new way of life*, London: Nicholson & Watson 1945.

Institut für Wissenschafts- und Technikforschung
Universität Bielefeld
Postfach 100 131
D-33501 Bielefeld
Germany
nikolow@iwt.uni-bielefeld.de

REVIEWS

PAOLO PARRINI, WESLEY C. SALMON & MERRILEE H. SALMON (eds.), *Logical Empiricism: Historical & Contemporary Perspectives*. Pittsburgh: University of Pittsburgh Press, 2003.

This volume presents seventeen essays (not eleven, as the publisher inexplicably claims) by a diverse group of philosophers that arose out of a conference in Florence in 1999. As its title indicates, the focus of the conference was the contemporary significance of the topics, methods and innovations of the logical empiricists. This has led to a nicely balanced collection that combines careful historical study with an eye on current debates in the philosophy of science and mind.

After a brief introduction by Parrini and Wesley Salmon, we find four essays grouped together under the heading "Turning Points and Fundamental Controversies." In "A Turning Point in Philososphy: Carnap – Cassirer – Heidegger", Michael Friedman summarizes his *A Parting of the Ways* (Open Court, 2000) and focuses on the details of the relationship between Carnap and Cassirer after the latter's break with neo-Kantianism. Friedman argues that the 1929 Davos debate between Cassirer and Heidegger on Kant, which Carnap attended, marks a crucial break in the history of philosophy. Both politically and philosophically, Cassirer's Marburg neo-Kantianism occupied a middle group between the more extreme programs of Heidegger and Carnap. Thus Cassirer's "astonishingly comprehensive body of philosophical work" (29) is presented as an unexplored attempt to fruitfully combine the scientific orientation of Carnap with Heidegger's humanistic ambitions.

Gottfried Gabriel's "Carnap's 'Elimination of Metaphysics through Logical Analysis of Language': A Retrospective Consideration of the Relationship between Continental and Analytic Philosophy" approaches the analytic-continental divide somewhat differently by emphasizing the similarities between Carnap and Heidegger as well as the influence of Dilthey's life philosophy on both. Gabriel points to the shared rejection of traditional metaphysics and psychologism but claims that Carnap's preference for a scientific orientation led him to dismiss Heidegger's more poetic preoccupations. He may go too far in concluding, though, "that the difference between Continental and analytic philosophy is above all a matter of the style of thought that manifests [itself] in linguistic style" (38).

An earlier dispute between analytic and continental philosophers is the focus of Roberta Lanfredini's "Schlick and Husserl on the Essence of Knowledge".

While for Schlick the qualitative aspects of our experience were something to be overcome in rigorous, scientific thinking, Husserl insisted that "authentic knowledge cannot do without a qualitative factor" (50). Siding largely with Husserl, Lanfredini notes the many inaccuracies in Schlick's criticisms of phenomenology and the resources that Husserl had to respond.

"Carnap vs. Gödel on Syntax and Tolerance" by S. Awodey and A. W. Carus rounds out part one, although their essay focuses on a dispute within the analytic tradition. They argue that Gödel's criticism of Carnap's approach to mathematics in *The Logical Syntax of Language* "contains a mistake" (57). The mistake was to augment the reasonable demand that any language that Carnap adopt for mathematics actually be consistent with the requirement that the language be *provably* consistent. The stronger requirement cannot be met, but Awodey and Carus claim that Carnap did not need to satisfy it.

Part two, "On the Origins and Development of the Vienna Circle", opens with Thomas Uebel's important essay "On the Austrian Roots of Logical Empiricism: The Case of the First Vienna Circle". Building on the results of his *Vernunftkritik und Wissenschaft: Otto Neurath und der erste Wiener Kreis* (Springer, 2000), Uebel describes a distinctively Austrian orientation in the philosophy of science that predates World War I. He traces its formation from the influential high-school textbook by Alois Höfler (a student of Brentano and Meinong) through to the series of reviews by Hahn and Frank for the *Monatshefte für Mathematik und Physik*. Of particular interest is Uebel's suggestion that the attitude towards Kant that Hahn, Frank and Neurath inherited was partly responsible for the later serious divisions in the 'second' Vienna Circle.

George Reisch's "On the *International Encyclopedia*, the Neurath-Carnap Disputes, and the Second World War" places the well known philosophical disputes between Neurath and Carnap in a broader intellectual and political context. Neurath sought to turn the Unity of Science movement and its *Enclyclopedia* into an international social movement, and this clashed with Carnap's more modest academic ambitions. Reisch relates how things came to a head when Carnap declined to edit Neurath's 1944 "Foundations of the Social Sciences". This led Neurath to accuse Carnap, in a letter from 1945 that was not sent, of having "many serious signs of a Platonic attitude" (103) that risked a slide from semantics to a kind of totalitarianism. Neurath's death later in 1945 stopped this dispute from escalating further.

Part two ends with Gereon Wolters' "C. G. Hempel: Pragmatic Empiricist", a defense of Hempel against some standard criticisms. Wolters claims that by the mid-1960s Hempel had come to recognize that pragmatic factors were needed to complete his proposals on confirmation and explanation. This pragmatic turn culminated in Hempel's surprising endorsement of Neurath in a 1991 address in Konstanz: "it is important that we look at Neurath and then say that it is his ideas that despite their somewhat primitive form give us hints [as] to how the philosophy of science has to be developed further" (117).

Part three, "The Riddle of Wittgenstein", contains only one essay: "The Methods of the *Tractatus*: Beyond Positivism and Metaphysics?" by David Stern. Stern masterfully sets out the history of *Tractatus* interpretation and argues that the lack of agreement is a sign that the text itself is not coherent. The more recent 'therapeutic' reading by Diamond and others is rejected based on Wittgenstein's own doctrinal approach to the text at the time. But instead of using this information to argue for another standard interpretation, Stern focuses on the conflicting philosophical methods that Wittgenstein employed. The paper concludes with a comparison of these interpretations of the *Tractatus* with debates in sociology about the *Philosophical Investigations*.

Thomas Ryckman's challenging "Two Roads from Kant: Cassirer, Reichenbach, and General Relativity" opens part four, "Philosophy of Physics". He argues that it was in fact Cassirer who "produced a significantly more insightful appreciation of the epistemological innovation of the theory of general relativity" (187) when compared to Reichenbach. Thus, like Friedman, Ryckman sees Cassirer as an important option that has been overlooked in contemporary debates. But, breaking somewhat with Friedman, Ryckman bases his conclusion on an interpretation of general relativity theory that emphasizes the connections between the local symmetries of gauge field theories and general covariance. One is left eager for more details about how these later mathematical developments relate to Cassirer's philosophy, and fortunately Ryckman's *Reign of Relativity* (Oxford, forthcoming) will supply this and more.

In "Vienna Indeterminism II: From Exner's Synthesis to Frank and von Mises", Michael Stöltzner builds on his earlier work on Mach, Boltzmann and Exner in fleshing out a picture of a distinctively Viennese openness to indeterministic physical theories, long before quantum mechanics. Here he relates how Exner influenced Frank and von Mises. At the heart of this approach was Mach's functional conception of causal laws, which "made it viable to contemplate indeterminism at the most basic level of reality" (196). Opposed to this was the more traditional Kantian conception of causality, here identified with Planck, which equated causality with determinism. The frequentism of Frank and von Mises is thus skillfully traced to its Viennese origins. Like Uebel's work, Stöltzner's discoveries allow a more nuanced understanding of later debates within the broader logical empiricist movement.

Part five, "The Mind-Body Problem", contains Michael Heidelberger's "The Mind-Body Problem in the Origin of Logical Empiricism: Herbert Feigl and Psychophysical Parallelism" and Jaegwon Kim's "Logical Positivism and the Mind-Body Problem". Despite Heidelberger's criticisms of Kim's earlier work as reflecting a "widespread (mis)conception of logical empiricism" (233), in this volume Kim offers a careful and fairly sympathetic review of the work of Schlick, Carnap and Hempel on the mind-body problem. Kim even grants that he "was impressed by the metaphysical depth and sophistication in the positivist philosophers, especially Carnap" (277). Heidelberger reconnects Feigl's identity theory with the nineteenth century German-language tradition that stretches from

Fechner, Lange and Riehl to Schlick and Carnap. In addition, he rather convincingly argues that these debates decisively shaped logical empiricism, e.g. in the rejection of explanation as a goal of science.

"Kinds of Probabalism" by Maria Carla Galavotti is the first of the three essays in part six, "Scientific Rationality". She contrasts the varying approaches to probability by Reichenbach, Carnap and de Finetti, and argues that Reichenbach's and Carnap's views are more similar than it might initially appear. Showing some sympathy for subjectivism, Galavotti concludes that de Finetti and Ramsey laid the foundation for "an interpretation of probability in tune with a pragmatist and nonrealist epistemology" (298).

Martin Carrier's "Smooth Lines of Confirmation Theory: Carnap, Hempel, and the Moderns" focuses on the roots of the 'modern' approaches to confirmation by Glymour and others in the earlier work of the logical empiricists. He insists that "Confirmation theory can be regarded as a thoroughly cumulative endeavor" (306) and supports this claim by showing how Glymour's 'bootstrap' model arose out of the problems with Hempel's account. 'Cumulative' here is given a rather weak reading, though, for as Carrier recognizes almost all of the central principles of Hempel's proposal were rejected by Glymour.

In "Changing Conceptions of Rationality: From Logical Empiricism to Postpositivism", Gürol Irzik defends both Carnap and Kuhn from the accusation that they are relativists. At the heart of his defense is an interpretation that allows Carnap to rationally prefer certain answers to external questions. This expanded conception of rationality is then linked to Kuhn's post-*Structure* writings on fixed scientific values. In the end, though, Carnap's conception of rationality fails because of his emotivist approach to value judgements.

The volume concludes with two essays on "Non-linguistic Empiricism": Paolo Parrini's "Reason and Perception: In Defense of a Nonlinguistic Version of Empiricism" and Wesley Salmon's "Commit It Then to the Flames ...". Parrini argues that empiricism is best served by the naturalistic project of analyzing the psychological facts that underlie our ability to use language. This prelinguistic focus is linked to Mary Hesse's network model of confirmation. Hesse assumes a primitive ability to recognize similarities in nature, and Parrini claims that this should be at the center of an empiricist epistemology.

Salmon reviews the Ayer-Church and Carnap-Kaplan disputes with the goal of determining a viable criterion of cognitive significance. These exchanges point the way to a non-linguistic and non-deductive form of empiricism that sees a weak form of verificationism as a kind of intellectual responsibility. Thus, Salmon concludes, "Empiricism is a viable philosophy that can stand on its own outside of the purely linguistic framework in which it was notoriously placed, as long as the critical role of nondemonstrative reasoning is clearly recognized" (386).

The wide range of topics, and the international makeup of the contributors, demonstrate the continuing appeal and importance of the writings of the logical empiricists. I was most struck by the attempts of contributors as varied as Fried-

man, Ryckman, Parrini and Salmon to show that contemporary philosophy has missed the best or most promising aspects of logical empiricism. The dangers of this approach are made manifest by the conflict between Salmon, who seems to model his non-linguistic empiricism on Reichenbach, and Friedman and Ryckman, who aim to discredit this very approach to science. As a historian, I worry that these debates about the value of a philosopher's work might become overly entangled with the question of what a philosopher actually thought. The current state of Frege and Wittgenstein scholarship remind us that it is, after all, difficult enough to come up with an adequate interpretation of a philosopher's writings and that appeals to what we now 'know' about philosophy are not likely to help. My hope, then, is that historians of logical empiricism will continue their focus on the careful reconstruction of the context and views of real historical positions.

Christopher Pincock

C. WAYNE MAYHALL, *On Logical Positivism*, Wadsworth Philosophical Topics, Belmont, 2003.

This book is an exercise in basic introductory work. Nothing more. C. Wayne Mayhall makes clear from the very beginning that his study is not intended to be a historical analysis of the roots and the internal development(s) of logical positivism. Rather, its aim is to "immerse the reader in the writings of the original members of the Vienna Circle" (p. i). In attempting to achieve this goal, however, Mayhall, in the final analysis, becomes a victim of his largely ahistoric method: neither the originality nor the significance of the Vienna Circle members' contributions to philosophy (and philosophical style) are clarified in a satisfactory manner.

The book consists of 14 short chapters, each of which is accompanied by a set of (for the most part) useful questions for discussion. Chapter 1, the Introduction, outlines the formative context of the logical positivist movement. According to Mayhall, the members of the Vienna Circle were influenced by two major contributions in the area of logic and logical analysis: Russell and Whitehead's *Principia Mathematica* (1910-1913) and Wittgenstein's *Tractatus Logico-Philosophicus* (1922). This is surely correct. But what about Mach, Einstein, Poincaré and, in particular, the new situation in physics around 1920? It is well known that the three main figures of the logical-positivist movement – namely Schlick, Reichenbach and Carnap – all began their philosophical careers with extended studies of issues raised by Einstein's Theory of Relativity. Mach's critique of Newtonian mechanics and Poincaré's advocacy of a conventionalism in geometry served as important reference points in this connection. Especially

the logical positivists' rejection of metaphysics must be seen as a result of their early preoccupation with the writings of these authors. Unfortunately, Mayhall does not even allude to these influences.

This omission in turn has consequences for chapters 2, 3 and 4, which deal with the logical positivists' attitude toward traditional philosophy and their alternative to it. To be sure, Mayhall clearly recognizes the anti-metaphysical impetus by which the members of the Vienna Circle were driven. But the reason they rejected metaphysics is not made sufficiently clear. On Mayhall's view, it was "the insight into to the nature of logic itself" (p. 20) that led the logical positivists to propagate a "turning point" (Schlick) in the history of philosophy. So far, so good. But what exactly was the context of their attempt to overcome traditional metaphysics and its contemporary varieties? Mayhall should at least have mentioned the significance of Heidegger's 'philosophy of being', especially for Carnap's refutation of – overtly *a-logical* – metaphysical thinking. However, there was a second – "hidden" (p. 30) – form of metaphysics that attracted the logical positivists' critique: Kant's transcendentalism and its conception of synthetic a priori truths. Mayhall correctly points out that the denial of the existence of such truths lay at the very bottom of the logical positivists' strictly empiricist account of scientific knowledge. But instead of making clear the central arguments they raised against the transcendental point of view, he gives a rather lengthy survey of the original Kantian doctrine (pp. 30-32). It would have been more helpful if he had focused on the logical positivist critique and on its anti-Kantian roots in Poincaré's conventionalism.

Chapters 5, 6 and 7 are devoted to the "cornerstone doctrine" (p. 51) of the Vienna Circle: the Principle of Verifiability and the associated empiricist criterion of meaning. Proceeding from Schlick's (or Wittgenstein's) famous dictum that the meaning of a statement is the method of its verification, Mayhall outlines, in a clear and sober way, both the implications of this view and its central problems. He elucidates the various strategies by which the members of the Circle gradually softened the Principle of Verifiability in terms of weaker principles like "testability", "confirmabililty" and "translatability" (pp. 63-72). Furthermore, the conception of protocol sentences is discussed at some length (pp. 73-83). As is well known, the central idea behind this conception was the vision of a unified language of science. But the epistemic status of protocol sentences was highly controversial. Schlick on the one hand and Neurath on the other were the main antagonists in this debate within the Vienna Circle. Mayhall, however, ignores the controversial aspects and instead confines himself to Neurath's point of view. This is legitimate but is nevertheless a pity, for the protocol sentence debate is a striking example for the often neglected – if not denied – *diversity of standpoints* within the Vienna Circle.

Chapter 8 is a good one. Here, the reader can learn a lot about Carnap's *Logical Syntax of Language* (1934) and some of its complexities. In a nutshell, the *Logical Syntax* was an attempt to reformulate the logical-positivist doctrine by means of a "formal theory" (p. 85). Mayhall summarizes the central pillars of

this theory in a thoroughly appealing manner. He uses various diagrams to help acquaint the reader with Carnap's basic insights concerning the aims and scope of a syntactical analysis of scientific language.

Chapter 9 starts from the opposition of the two most prevalent theories of truth: the correspondence theory of truth and the coherence theory of truth. According to Mayhall, the members of the Vienna Circle converted one after the other from the first to the second theory (p. 99). So much, in any case, is testified by Carl Gustav Hempel, who held this view in an influential article in 1935. Nevertheless, one must be cautious when it comes to the details of this appraisal. Schlick, for example, *never* advocated a coherence theory of truth. And in the case of Carnap it is interesting to note that he – at the famous Paris Congress in 1935! – defended *Tarski's theory of truth*, the aim of which was to semantically explicate the correspondence view of truth. In short: Mayhall's reliance on the story told by Hempel is, to say the least, historically problematic.

Chapters 10 and 11 deal with the less well-known aspects of the logical positivist doctrine: its application to the areas of ethics and sociology. As for the first, Mayhall gives a useful summary of Schlick's *Fragen der Ethik* (1930). As for the second, he offers an interesting introduction to Neurath's "Soziologie im Physikalismus" (1931/2). Of particular interest are the parallels he draws between Neurath's approach to sociology and the one put forward by Karl Marx. According to Mayhall, the significance of Neurath's adoption of a Marxist theory of sociological laws lay in the "elimination of meaningless moral science" and in "the exaltation of sociology as natural science" (p. 122). It was these two aspects which, in Neurath's view, spoke in favor of Marxism as the historically best example available of an empirical sociology that could be used for the project of unifying science in terms of physicalism and behaviorism.

Forget about the last three chapters. They are confusing and rather weak in terms of content. Here is an example: Wittgenstein's influence upon the members of the Vienna Circle (chap. 13) is reduced to his 'invention' of the truth-table method in first-order logic (pp. 139-146). It is really hard to believe that this was the reason why outstanding members of the Circle (esp. Schlick and Waismann) were so impressed by the approach put forward by the author of the *Tractatus*. It rather seems that Wittgenstein's impact had to do primarily with the afore-mentioned "insight into the nature of logic itself".

In summary, it must be said that Mayhall's contribution is a case of light and shadow; in places readable and useful, but in the final analysis too superficial and historically unrelated.

Matthias Neuber

C. WAYNE MAYHALL, *On Carnap*, Wadsworth Philosopher Series, Wadsworth / Thomson Learning, Belmont USA, 2002.

Mayhall's book is intended as a short introduction to the philosophy of Rudolf Carnap. It is a collection of easy-to-grasp, but completely ahistoric condensations of some, more or less, crucial examples of Carnap's work. Unfortunately those examples are all products of the short period between 1928 and 1932, which *is,* to be sure, crucial for the development of Carnap's ideas and for the development of Logical Empiricism as a whole. Nevertheless, it seems unmotivated to reduce even an introductory book to this period.

Apart from a short introduction, which provides a couple of useful quotations from Carnap's intellectual autobiography, Mayhall's book consists of six chapters that comment on the following texts of Carnap: "The Logical Structure of the world" (Chapter 2), "Logical Foundations of the Unity of Science" (Chapter 3), "The Old and the New Logic" (Chapter 4), "Pseudo-Problems in Philosophy" (Chapter 5), "The Elimination of Metaphysics Through Logical Analysis of Language" (Chapter 6) and "Psychology in Physical Language" (Chapter 7). The chapters draw heavily upon numerous quotations from the respective original texts; there are some more or less instructive examples illustrating Carnap's claims; no lengthy descriptions but numerous clearly arranged lists of core theses and definitions of basic concepts. Thus Mayhall's presentation is, on the whole, clear and concise, resembling a clear and concise textbook for class.

The problem of Mayhall's approach to the philosophy of Carnap is not so much the presentation, but rather that – although the book is more or less verbally reproduces Carnap's phrases – it is hard to find much about Carnap in it. What Mayhall is reproducing tends to be something like the *conceptual wrapping*, without the stuff inside. For example, Mayhall confronts us with a short introduction to 'logic as tautology' (Chapter 4), where he introduces some truth-table definitions of truth functors. We learn that 'the most important concepts of the new logic' are '(1) Negation: "not", (2) The logical connectives for two sentences: "and," "or," "if-then", (3) "every" (or "all"), "there is", (4) "identical"' (p. 50). It is true, those concepts are important, but, in which sense are they 'concepts' and how can we use them? Mayhall does not provide an answer. He says *nothing* about quantification and identity, let alone that he provides a couple of examples from natural languages. It is, to be sure, not indispensible to give an introduction to logic in an introductory book about Carnap, but *if* one tries to give such an introduction to formal concepts – and Mayhall does – then one has to clarify them. We do not find such clarifications. And that's the way the whole book is written: as a collection of argument and concepts out of Carnap's work and out of modern logics and mathematics, without any *useful* explanation and clarification, especially for a student reader doing his or her first steps in Logical Empiricism.

Nevertheless, Mayhall's method could be acceptable, taking the book as an inexpensive way to obtain some snapshots of Carnap's work. But even in that case it is hard to recommend the book because of its restriction to the period between 1928 and 1932. – What about the semantical turn from logical syntax to Tarski-style semantics? What about the 'Logical Syntax of Language' and 'Meaning and Necessity', as the second and third of Carnap's main contributions? What about such an influential, and easy-to-read, paper such as 'Empiricism, Semantics and Ontology'? What about Carnap's, more difficult, later work on probability and induction? And finally: what about Carnap's intellectual relations, from Russell, Frege and Wittgenstein, through Neurath and the whole Vienna Circle until Quine and the Anglo-American scene? We do not even find random quotations from the fields dealt with in Mayhall's book.

In summary, it is hard to believe that Mayhall's approach could provide the typically student readership with a trustworthy impression of Carnap's work.

Christian Damböck

JEAN-PIERRE COMETTI and KEVIN MULLIGAN, *La philosophie autrichienne de Bolzano à Musil. Histoire et actualité*, Paris: J. Vrin, 2001.

Bolzano and Musil conveniently mark the bounds of this collective volume on Austrian philosophy. The former published in 1804, as the 19th century began, his first article, which dealt with mathematics; Bolzano was to move on to philosophy, becoming the initiator of a highly original and fruitful tradition. Musil, passing away during World War II, marks the end of a speculative experience passed on from one generation to the next. What had made Vienna and *Mitteleuropa* so conducive to innovation was then lost. *La philosophie autrichienne de Bolzano à Musil* is the outcome of a colloquium held at the International Cultural Center of Cerisy la Salle in 1997. It is a sequel to a special issue on the critique of reason in Central Europe, which appeared in the journal *Philosophiques*. This second volume allows for an in depth analysis and an integrated presentation. The thirteen articles are presented under five headings related to modern areas of investigation, providing a panorama of Austrian philosophy during the century and a half under study: 1. Metaphysics and logic, 2. Science, mathematics and epistemology, 3. Mind, 4. Values and norms, 5. Society.

A pungent introduction by K. Mulligan opens the volume. He defends the value of the Austrian school, the matrix of 20th century philosophy, criticizes French-speaking philosophers for their neglect of this tradition and presents the results of recent scholarship in this area. While giving an overview of the main

themes of the volume, Mulligan provides the reader with several stimulating suggestions concerning the nature of what could be termed the Austrian miracle. The following two articles offer a helpful cross-section of Austrian thought. P. Simons describes the development of the theory of being from Bolzano to the early Wittgenstein. He gives tools for evaluating the diverse ontologies formulated and help us to understand what lies behind the anti-metaphysical stance to emerge in the 20th century. J. Wolenski, in turn, provides a thorough investigation of theories of truth. Following up on the history of these theories, he is able to show that the Polish school was not unconnected with Austrian philosophy. This is particularly illuminating with respect to Tarski, whose masterly achievement tends to conceal the origins of his thought.

The subsequent articles focus on individual thinkers or the relation between two thinkers: Boltzmann, Bolzano, Ehrenfels, Hayek, Meinong, Carl Menger and Musil. Among these Meinong receives the most attention. He is studied with respect to his conceptions of modality and reference, his epistemology as well as his theory of emotions. This reveals a complete change of attitude: after being subjected to the criticism by no lesser than Russell, Husserl and Brentano, Meinong's bountiful ontology and his endeavors in ethics have provoked new interest. A philosopher should be judged by the value of the debates he initiated, as Nef puts it, who submits Meinong's modal theory to a careful analysis. He goes on to add: "Meinong illustrates the main features that are commonly ascribed to Austrian philosophy – empiricism, openness to psychology, inclination for description and detail, recourse to conceptual analysis – as distinguished from German philosophy – speculative idealism, anti-psychologism, predilection for encompassing systems" (p. 81).

Wittgenstein, Popper and the members of the Vienna Circle are mentioned only as concerns their relationship to the above-mentioned thinkers. The aim of this book is to impart knowledge of less well-known figures of the Austrian tradition, the notable exception being Husserl. But this is not surprising in view of ongoing research on Husserl as a pupil of Brentano. J. Dokic proposes a close study of the latter's 1907 lectures *Ding und Raum*. He develops a comparison with Poincaré's genesis of our notion of space, thereby bringing out the implicit criterion of reversibility underlying Husserl's attempt. Husserl is studied by M. Konrad on an equal footing with Meinong as regards ethics and supererogation. The commentator notes that Austrian thinkers depart here from Kant; the classification of moral acts is richer. He emphasizes the realistic trend of Austrian philosophy, the insistence on a complete and careful analysis of real facts. J. Bouveresse gives a lucid account of the state of Boltzmann studies as well as stimulating guidelines for future research. In the past decade important material has surfaced: the belated publication of his *Principien der Naturphilosophie* and the discovery of a *Nachlass*. Against the tendency among Vienna Circle members to minimize the differences between Boltzmann and Mach, Bouveresse brings out the specificity of the position of the former with respect to realism. He also shows how Boltzmann's influence intricately combined with that of Hertz in

the case of Wittgenstein. P. Rusnock addresses the connection between Bolzano's mathematics and his philosophy. We are shown why Bolzano, who is a highly creative mathematician parted paths with Kant. Mathematics since the middle of the 18th century had taken an altogether new direction and was no longer suitably described by the Kantian notion of construction. Bolzano was attentive to the accomplishments of Euler and Lagrange, and he sought to give them a philosophical basis. As Rusnock judiciously observes, "One perceives in the Bolzanian reconstruction of Lagrange's analysis the first conscious, systematic and enduring application of what will be later termed analytic philosophy" (p. 137).

Mulligan develops a line of criticism, to be substantiated in this book, that is directed against the long and unjust neglect of the Austrian tradition and analytic philosophy in France. Not only French thinkers have thereby neglected a large part of contemporary philosophy, but they have also ignored the origins of their own tradition. This accusation is leveled at Sartre and Foucault in particular. Phenomenology underwent a profound transformation, shifting from realist to idealist and transcendental. This change, as Mulligan points out, is still not fully understood (p. 23). This prompts me to voice a regret. There are some indications of links between Austria and France: Bolzano was acquainted with Lagrange and his school of mathematical analysis; similarities are noted between the contemporaneous theories of Husserl and Poincaré on spatial perception. As this book seeks to fill in a gap in French scholarship, it would have been interesting to locate this undertaking itself. G.G. Granger's *Ludwig Wittgenstein,* published in 1969, is commonly credited as having been the first book in French on the Austrian philosopher, which is surely a tardy reception. But a change has taken place since, and the research presented in this volume in fact builds on previous studies, which have increased significantly in the past few years.

To say that Austrian thought has taken time to be recognized in France is not to tell the whole story. Obviously, Husserl was well received, and it would be illuminating to contrast precisely those thinkers who were accepted and those who were banished. Moreover, Mulligan tells us that Couturat was aware of Bolzano's mathematical conceptions as well as the work of his student Zimmermann. Reference is also made to a publication by Arthur Hannequin in 1895, which unfortunately slipped out of the bibliography. This is his noteworthy dissertation *Essai critique sur l'hypothèse des atomes dans la science contemporaine*, in which he discusses the critical presentation of Cantor's theory given by Benno Kenny, a student of Brentano. Likewise, Bouveresse recalls that Renouvier was classed by Brentano as a negator of the continuity of space and time, alongside Boltzmann. One could add the debates stirred up at the time by Mach's views. Those who were engaged in France in developing philosophy of science appear to have been curious about what was happening in Austria; they were likewise forgotten by their successors. Reception is indeed a complex phenomenon and necessitates a careful historical study. Some fruitful exchanges thus took place at the turn of the 19th and 20th centuries, which then got checked. What

brought about this subsequent parting of ways? This question has yet to be fully answered. But it should be granted that Cometti and Mulligan have provided French scholarship with a valuable volume on major topics in Austrian philosophy and have brought together a distinguished group of international scholars.

Anastasios Brenner

STEVEN BELLER (ed.), *Rethinking Vienna 1900.* Berghahn Books, New York-Oxford 2001.

The subject of rethinking in this volume is mostly Carl E. Schorske's *Fin-de-Siècle Vienna. Politics and Culture*, the book for which its author received the Pulitzer Prize. In a learned introduction to the essays included in the book under review Beller takes us both into "the Life and Times of Vienna 1900", and to the texts that have dealt with those "Life and Times", and especially to those which have been concerned with them after Schorske had published his important work whose thesis was that the culture of the period under discussion was due to the lack of participation of the German middle class in the political process. This lack of participation was due to the unwillingness of the aristocracy to permit the middle class of the monarchy to govern. An important point Beller makes is that the German middle class referred to by Schorske consisted mostly of Jews. There were exceptions however: Carnap and Schlick in philosophy, Renner and Schumpeter in politics and economics. In architecture and painting the most significant contributions were made by non-Jews, and in literature and music some of the most important contributions were made by non-Jews. In this context of course a Jew is defined by whether or not his parents, and/or his grandparents were Jewish because most of them had left the Jewish religious community.

Beller's "Introduction" should be read together with Mary Gluck's "Afterthoughts". Gluck not only reintroduces Schorske's contribution; she points to, and stresses his history of the dehistorization of the modernism of various disciplines, to the failure of liberalism, and the emergence of mass politics on the one hand, and aestheticism on the other.

The included essays frequently – but not always – refer to Schorske. Janik asks for a more differentiated way of looking at Fin-de-Siècle Vienna, for models of explanation other than Schorske's. Judson points out that the Socialists were less liberal than Schorske had assumed, and that they had engaged in a cultural politics like the Christian Socials and the Nationalists. Shedel treats the Austrian "Sonderweg" which failed to follow – like Germany – the course to modernity. Burri points to the importance of an aristocracy for the way of life for

the bourgeoisie. Spector criticizes the exclusive use of dichotomies in Schorske's historical explanations. Pfabigan discusses "the psychoanalytic view of the world", a view that succeeded more through popularizations of psychoanalysis than through its scientific community. Hacohen deals with the tension between Jewish identity and culture using – so to speak – Karl Popper's cosmopolitanism as a model, Jensen deals with the artist's professional identity, using criteria that had been supplied by Schorske, and Sarmany-Parsons takes to task the projection of the Viennese pattern onto the whole of Austria-Hungary, and turns our attention to Budapest, Prague and Cracow.

This text should play a significant role in the ongoing discussion of the cultural and intellectual history of Austria-Hungary. It is important because it exposes some of the weaknesses of Schorske's work and of its main arguments. Without Schorske, of course, these critical contributions could not have been made. We must await the results of further argumentation, deliberation, and research.

Kurt Rudolf Fischer

ECKART MENZLER-TROTT, *Gentzens Problem. Mathematische Logik im nationalsozialistischen Deutschland*, Birkhäuser: Basel-Boston-Berlin 2001.

The author (EMT) tackles an outstanding task in the history of proof theory and foundations of mathematics, i.e., writing a comprehensive biography about the logician and proof theoretician Gerhard Gentzen (1909-1945) – the inventor of Gentzen type calculi of natural deduction and the sequent calculus. According to Solomon Feferman, Gentzen's work on structural and ordinal proof theory was most influential for the development of modern reductive proof theory in practice. When it comes to books dealing with Gentzen, there has so far only been the edition and publication of the *Collected Papers of Gerhard Gentzen* (1969) by Manfred Szabo (including a *Biographical Sketch* and the achievement-related *Introduction*). EMT's book is hardly comparable to the bulk of texts treating Gentzen-related issues. Unfortunately, neither Gentzen's achievements nor the (further) developments of his ideas are treated appropriately. To be sure, with respect to the size (425 pages, of which the plain text amounts to roughly 300 pages) of EMT's book, this may be rather considered a separate task, though the included paper by Jan von Plato about Gentzen and structural proof theory (*Gentzen und die Beweistheorie*) deals with an outline on one of Gentzen's topics. The book also contains three popular lectures by Gentzen about *The Concept of Infinity in Mathematics* (*Der Unendlichkeitsbegriff in der Mathe-*

matik), *The Concept of Infinity and Consistency of Mathematics* (*Unendlichkeitsbegriff und Widerspruchsfreiheit der Mathematik*), and *The Present State of Research into the Foundations of Mathematics* (*Die gegenwärtige Lage in der mathematischen Grundlagenforschung*) as well as four Gentzen-related timetables and 22 hitherto mostly unknown pictures.

In the preface, EMT sets himself two aims: to inform (i) in detail about Gentzen's life and (ii) about the situation of mathematical logic during the time between 1933 and 1945. The chapters 1 to 3 as well as 5 and 6 are devoted to (i) while chapter 4 is dedicated to (ii). Additionally, EMT voices his support for a publication of the Collected Works of Gentzen in German (p. XVII).

Chapter 1 deals with Gentzen's life until the emergence of the program of his Ph.D. thesis in 1932. This part includes a lot of new and so far unpublished sources about Gentzen's family, his youth and studies. Particularly remarkable are the poems by Gentzen at the age of seven and fifteen and the lucid documents about the schoolboy Gentzen.

In chapter 2, EMT's statement expressed in the preface (p. XI), namely that this book is not about proof theory or its history, proves to be meant seriously. If the reader expects something in detail about the development of Gentzen's scientific achievements, he will be disappointed. The same goes for chapters 3 and 5. Like in all other chapters, a lot of quotations form a long chain winding throughout the whole text. E.g. if one is interested in the circumstances concerning the unpublished original Gentzen consistency proof for number theory from 1935, one will find only some uncommented letters to Bernays (p. 52 and p. 58f.), a review by Gentzen written during that time, letters from and to Gentzen regarding his assistant job at Göttingen (p. 53f.), but no information in detail, e.g. about the "fan theorem" (the use of which by Gentzen EMT only hints on via a quotation on p. 63). On the other hand, there are a lot of quoted reviews and hitherto unpublished letters concerning the first (published) consistency proof for number theory. Regarding the scientifically relevant chapters 2, 3 and 5, EMT's treatment is particularly sad because Gentzen's scientific biography gets reduced to a set of quotations. This is not in line with the standards of scientific historiography. Despite EMT's "disclaimer" on p. XI, some explanatory remarks on Gentzen's achievements really would have been in order.

Chapter 3 deals with Gentzen's military service (1939-42) and the time of his "high reputation". This part contains a lot of quoted reviews written by Gentzen along with reviews about his published works. EMT concludes this chapter with a highly speculative account of Gentzen's release from military service due to a "state of nervous exhaustion" (p. 113). It also features (like all other chapters) breaks in the narrative, e.g. "Gentzen reads Tarski" – [quotation] – something about Gentzen's assistant job at Göttingen (p. 91); "Gödel left Göttingen and arrived at Berlin" – "Gödel lectures at Yale" (p. 100) and so on.

Chapter 4 is by far the largest one (130 pp.) and dedicated to (ii). With respect to the other chapters this part looks very isolated, but considering its emphatic character it might have been the central motive for EMT to write his

book. The chapter focuses on two topics: (iii) the (ab-)use of Gentzen by the Nazis as an advocate for their racist ideology and (iv) the relationship between scientific practise and political administration (p. 119). Regarding (iii), especially Friedrich Requard abuses Gentzen's contributions to the foundations and philosophy of mathematics in an ideological way (without Gentzen's agreement and knowledge). In devising a racist "foundation" of "German mathematics", Requard uses Gentzen's clarifying discussion of the distinction between the "potential infinite" and the "actual infinite" as an argument for his racially thesis that there are distinct forms of mathematical thinking (ch. 4K). In a similar vein, Max Steck interprets Hermann Weyl's article *The mathematical way of thinking*. EMT stresses (p. 200) that besides Gentzen the emigrants Gödel and Weyl were abused as witnesses for "German mathematics" as well. Gentzen himself was neither a collaborator of the NS, nor a victim (he had no considerable disadvantages). Regarding (iv), EMT states that the so-called "external and objective historiography" (EOH) only tells the stories of the emigrants and the downfall of mathematics in Germany during that time, while the names and actions of the mathematicians who stayed in Germany are barely mentioned as they were not even victims but merely opportunists (*[d]ie Namen und Taten der Zurückgebliebenen tauchen nur als kurze Erwähnung auf, denn sie sind nicht einmal Opfer, sondern nur Mitläufer*) (p. 142). EOH allegedly treats those who stayed either not at all or as related to the evil of National Socialism. EMT writes (p. 148): "The external and objective historiography would like to deny that the men [i.e. the mathematicians who stayed in Germany, M.W.] could at all further practice mathematics or that this would be mathematics in the tradition of Klein and Hilbert. But I claim exactly this." (*Die externe und objektive Geschichtsschreibung würde gerne ableugnen, daß diese Menschen* [the mathematicians in Germany, M.W.] *überhaupt noch Mathematik machen konnten oder das dies Mathematik im Sinne der Tradition von Klein und Hilbert war. Aber genau dies behaupte ich.*) This might be considered the central argument of the book. Accordingly, EMT focuses not on the emigrants, but mainly on mathematicians who stayed in Germany. If we think of the extensive treatment of prominent emigrants like Carnap, Gödel or Weyl in historiography, the scope of this chapter may seem to provide a welcome counterpart, but it fails due to the inadequate treatment and to its false basic assumption. For in fact no historiographer holds the position EMT ascribes to EOH. Instead, EMT burns some straw-men (i.e. Steck, May and Dingler) – only suitable biographical aspects are used in constructing the personalities of these "opponents", e.g. the lucid works by Gereon Wolters on Dingler's biography during the Nazi period do not even get mentioned. EMT concludes that the majority of the mathematicians who stayed in Germany were neither NS-collaborators, nor opportunists; nor had mathematics in Germany 'died' – a result far less earth-shattering than the author insinuates since no one had really ever claimed the opposite! Regarding (iv), according to EMT, mathematical logic did nor represent a danger to the Nazis because of its character (p. 160). A characteristic feature of this chapter is the author's style. Emphatically and not

shy in choosing his words, EMT sketches the political activities e.g. of Max Steck or Eduard May. In this context, some remarks are particularly strange, e.g. the "mania of foundationalism" [...] of Dingler and his epigons ("Fundierungswahn" [...] von Dingler und seinen Epigonen) (p. 174), Dingler's and Steck's cravings for elimination [of opponents, M.W.] (*Die Liquidationsgelüste von Dingler und Steck*) (p. 194), or that Dingler issues Thüring a denazification certificate (*Persil-Schein*) (p. 209).

Chapter 5 is devoted to Gentzen's move to Prague, his job in solving differential equations as a part of the V2 (rocket) project and also deals with the question of why Gentzen didn't leave Prague when it got dangerous even though he had opportunities to do so.

Chapter 6 focuses on the last few months in Gentzen's life and on his dramatic death. Some of the material is already discussed by Premysl Vihan: *The Last Months of Gerhard Gentzen in Prague* (in: *Collegium Logicum* Vol. 1, 1995, pp. 1-7). The quoted letters convey compelling impressions about Gentzen's arrest on the seventh of May and the cruel treatment in jail. As for the question why Gentzen did not leave Prague in time the reports further support the hypothesis that Gentzen underestimated the danger since he was completely absorbed by scientific issues.

In sum, the book is dominated by quotes and footnotes. Some quotations (especially the reviews of Gentzen's works) can only be understood appropriately if one has knowledge of technical details. EMT does not bother to make any clarifying comments in this respect. A substantial part of the footnotes deals with irrelevant or confusing information. Overall, this book is best considered a pure source book on Gentzen (as such it is marvellous), but otherwise the manuscript was just not ready for publication since it reads like a file-card box. Even highly interested people will be bored with nearly every page, not because of the lucid and unknown quotations, but due to the presentation by EMT. In collecting all the quoted sources EMT did a time-consuming, diligent and enormous job. But when compared to other biographical works on logicians and/or proof theoreticians (e.g. *Logical Dilemmas: The Life and Work of Kurt Gödel* (1997) by John Dawson), it clearly shows that the contents should have been handled more carefully. We can only hope that the forthcoming English translation will at the same time comprise a thorough revision.

Matthias Wille

ACTIVITIES OF THE INSTITUTE VIENNA CIRCLE

ACTIVITIES 2003

Österreich und der Nationalsozialismus: Die Folgen für die wissenschaftliche und humanistische Bildung / Austria and National Socialism: Implications for Scientific and Humanistic Scholarship
Location: University of Vienna, Kleiner Festsaal
Date: June 5–6, 2003
Lecturers: Mitchell Ash, Jean-François Bergier, Evan Burr Bukey, Christian Fleck, Eveline Goodman-Thau, Max Haller, Friedrich Hirzebruch, Gerald Holton, Clemens Jabloner, Eric Kandel, Ruth Klüger, Raoul Kneucker, Walter Kohn, Harry Lustig, Helga Nowotny, Peter Pulzer, Karl Sigmund, Ruth Lewin Sime, Friedrich Stadler, Fritz Stern, Edward Timms, Anton Zeilinger
http://www.univie.ac.at/ivc/Kandel/

Third Vienna International Summer University
Scientific World Conceptions
"Biological and Cosmological Evolution"
organized by the University of Vienna and the Institute Vienna Circle
Location: Vienna, University Campus
Date: July 14–27, 2003
Main Lecturers: Karl Sigmund (University of Vienna, Austria), Eörs Szathmáry (Eötvös Loránd University, Hungary), Robert M. Wald (University of Chicago, USA)
Assistant Lecturer: Daniel Holz (University of California, Santa Barbara, USA)
http://www.univie.ac.at/ivc/VISU/

ESF Network: Historical and Contemporary Perspectives of Science in Europe
Third International Workshop, London: "Laws and Models in Science"
Location: King's College, London
Date: September 6–10, 2003
Lecturers and Commentators: Jeremy Butterfield (Oxford), Thomas Placek (Cracow), Tim Shallice (London), Eric Myin (Brussels), Dov Gabbay (London), Gerhard Schurz (Düsseldorf), Katarzyna Paprzycka (Poznan), Marcel Weber (Hannover), Mary Morgan (London), Peter Rosner (Vienna), Rainer Hegselmann (Bayreuth), Thomas Mormann (San Sebastián), Eric Weber (Gent), Pascal Engel (Paris), Wlodek Rabinowicz (Uppsala), Martin Carrier (Bielefeld), James

McAllister (Leiden), Maria Carla Galavotti (Bologna), Donald Gillies (London), How Price (Sydney/Edinburgh), Raffaella Campaner (Bologna)
http://www.esf.org/

Paris – Wien. Enzyklopädien im Vergleich / Paris – Vienna. A Comparison of Encyclopedias
Location: Institut Français, Palais Clam Gallas, Vienna
Date: October 16–18, 2003
Organised together with the Institut Français
Lecturers: Anastasios Brenner (Paris), Robert S. Cohen (Boston), Hans-Joachim Dahms (Berlin), Dominique Lecourt (Paris), Mathieu Marion (Montréal), Thomas Mormann (San Sebastián), Elisabeth Nemeth (Wien), Mélika Ouelbani (Tunis), George Reisch (Chicago), Peter Schöttler (Paris/Berlin), Friedrich Stadler (Wien), Antonia Soulez (Paris), Walter Tega (Bologna), Thomas E. Uebel (Manchester)
http://www.univie.ac.at/ivc/Encyclopedia/

Cambridge and Vienna. Frank P. Ramsey and the Vienna Circle
Location: Vienna, University Campus
Date: November 28/29, 2003
Lecturers: Patrick Suppes (Stanford), Michael Dummett (Oxford), Eckehart Köhler (Vienna), Wolfgang Degen (Erlangen), Peter Clark (St. Andrews), Brian Skyrms (UC Irvine), Stathis Psillos (Athens), Brian McGuinness (Siena), Gabriele Taylor (Oxford), Maria Carla Galavotti (Bologna)
http://www.univie.ac.at/ivc/Ramsey/

Scientific World Conception and Art
Art, Theory of Art and Studies in Art in the Scientific Discourse
Coordination: Martin Seiler
Presentations: Sabine Koch (Wien): Das Projekt der Ringbühne und seine Entwicklung bei Oskar Strnad: Theaterarchitektur – neue Sehweisen – Ostasienrezeption; Sebastian Meissl (Wien): Österreichische Gelehrtenbriefe des 19. Jahrhunderts: Der Nachlass des Germanisten Wilhelm Scherer (wissenschaftsgeschichtlicher Quellenwert, Probleme der Edition und Kommentierung); "KunstKunstKunst. Der große österreichische Staatspreis 1950-2002". Gespräch mit dem Kurator John Sailer im Anschluss an die Führung am 11. April durch die Ausstellung; Károly Kókai (Wien): Franz Wickhoff und die Geistes- und Kulturgeschichte des 19. Jahrhunderts.
http://www.univie.ac.at/ivc/WWUK

Moritz Schlick (1882–1936): Critical Complete Edition and Biography
The goals of this planned 10-year research project, financed by the Austrian Research Fund (FWF), which will begin with a three-year start-up phase, are the following: to put together a Critical Complete Edition of Moritz Schlick's Writings and to write a Biography (12 plus 1 volumes) within the context of the his-

tory of philosophy and science. The ongoing research in philosophy of science will help clarify still open questions and facilitate future studies in this area. The planned edition will for the first time make Schlick's entire work, which until now was only available in a fragmentary way, accessible to a large circle of students and scholars and interested readers. The edition will be edited and commented on in German.

This international research and edition project is being carried out as a joint initiative by the *Institute Vienna Circle* based in Vienna (A), the *Research and Documentation Center for Austrian Philosophy* (*Forschungsstelle und Dokumentationszentrum fuer Oesterreichische Philosophie*) in Graz (A) and the *Department of Philosophy* (*Institut fuer Philosophie*) at the University of Rostock (G). Moritz Schlick's writings (housed at the Rijksarchief in Haarlem, NL) that are crucial for this project have been archived and copied onto CD-ROM so that they are now accessible in Vienna, Graz and Rostock. The administrators of the Schlick papers have agreed to sign a publication contract with Springer Verlag (Vienna–New York). For the first time all (published and unpublished) writings including correspondence will be accessible to researchers and the public. Parallel to this, the first intellectual biography on Schlick will be put together.
http://www.moritz-schlick.de/

Publications

The Vienna Circle and Logical Empiricism. Re-evaluation and Future Perspectives
Ed. by Friedrich Stadler. Dordrecht–Boston–London: Kluwer, 2003 (Vienna Circle Institute Yearbook, vol. 10)

Language, Truth and Knowledge. Contributions to the Philosophy of Rudolf Carnap
Ed. by Thomas Bonk. Dordrecht–Boston–London: Kluwer, 2003 (Vienna Circle Institute Library, vol. 2)

Wissenschaftsphilosophie und Politik / Philosophy of Science and Politics
Ed. by Michael Heidelberger and Friedrich Stadler. Wien: Springer, 2003 (Veröffentlichungen des Instituts Wiener Kreis, Bd. 11)

Intellectual Migration and Cultural Transformation. Refugees from National Socialism in the English-speaking World
Ed. by Edward and Jon Hughes. Wien: Springer, 2003 (Veröffentlichungen des Instituts Wiener Kreis, Bd. 12)

Wien und der Wiener Kreis. Orte einer unvollendeten Moderne
Hg. von Volker Thurm-Nemeth unter Mitarbeit von Elisabeth Nemeth. Wien: WUV Verlag, 2003

Library and Documentation

Expansion of primary sources and research literature on the Vienna Circle and its influence.
Acquisition of estates and archival material in Austria and abroad.

Robert S. Cohen Collection and Archives: Robert Sonné Cohen (b. 1924) is an American philosopher, scientist and historian of science and philosophy who has been editing the *Boston Studies in the Philosophy of Science* and organizing the *Boston Colloquium for the Philosophy of Science* for many decades. The Robert S. Cohen Collection contains correspondence, unpublished and published manuscripts and typescripts, reprints, journal issues, news clippings, photographic prints, sound recordings, memos and notebooks.
A substantial portion of the Robert S. Cohen Collection is being made available for use by educators and researchers at the Institute Vienna Circle as the Robert S. Cohen Archives. The Institute Vienna Circle will provide access to full-sized photomechanical reproductions of documents selected from the Collection that offer insight into the development of Logical Empiricism.

Eugene T. Gadol Library: The research library of Eugene T. Gadol has been dedicated to the Institute Vienna Circle in 2002 by his heirs and is available for research.

PREVIEW 2004

Workshop
Sigmund Exner – Physiologie, Psychologie, Ästhetik und empirische Kulturforschung um 1900
Location: University of Vienna, Department of Contemporary History
Date: March 5/6, 2004
Organised together with the Max-Planck-Institut für Wissenschaftsgeschichte in Berlin
Lecturers: Mitchell Ash (Wien), Gerhard Benetka (Wien), Herta Blaukopf (Wien), Karl Clausberg (Lünebrug), Debbie Coen (Harvard), Peter Geimer (Berlin), Michael Hagner (Berlin), Anke te Heesen (Berlin), Veronika Hofer (Wien), Christoph Hoffmann (Berlin), Gerda Lechleitner (Wien), Burkhard Stangl (Wien), Michael Stöltzner (Bielefeld), Paul Ziche (München)

Fifth International Fellows Conference of the Center for Philosophy of Science
Recent Developments in the History and Philosophy of Science
Together with the Center for Philosophy of Science (Pittsburgh), Department of Philosophy (University of Cracow) and the Centre of Philosophy and Philosophy of Science (University of Konstanz)
Location: Cracow (Poland)
Date: May 26–29, 2004

Paul K. Feyerabend Revisited. On the Occasion of his 80th Birthday
Location: University of Vienna
Date: June 18/19, 2004

Fourth Vienna International Summer University
Scientific World Conceptions (VISU-SWC)
"The Quest for Objectivity"
Location: Vienna, University Campus
Date: July 19–30, 2004
Lecturers: John Beatty (University of Minnesota, USA), Michael Friedman (Stanford University, USA), Helen Longino (University of Minnesota, USA)

Publications

Induction and Deduction in the Sciences
Ed. by Friedrich Stadler. Dordrecht–Boston–London: Kluwer, 2004 (Vienna Circle Institute Yearbook, vol. 11)

Cambridge and Vienna. Frank P. Ramsey and the Vienna Circle
Ed. By Maria Carla Galavotti and Eckehart Köhler. Dordrecht–Boston–London: Kluwer, 2004 (Vienna Circle Institute Library, vol. 3)

Austria and National Socialism. Consequences for the Intellectual Life
Ed. by Friedrich Stadler. Wien: Springer, 2004 (Veröffentlichungen des Instituts Wiener Kreis, Bd. 13)

Paris – Wien. Enzyklopädien im Vergleich
Hrsg. von Elisabeth Nemeth. Wien: Springer, 2004 (Veröffentlichungen des Instituts Wiener Kreis, Bd. 14)

REMEMBERING DICK JEFFREY
(1926 – 2002)

On November 9, 2002, Richard Jeffrey died in Princeton at the age of 76. Jeffrey – Dick to all who knew him – was Emeritus Professor of Philosophy at Princeton University, where he obtained his Ph. D. in 1957. After a long struggle against cancer, which allowed him extended periods of well-being, Dick had to surrender, but before the final defeat he was able to finish a book called *Subjective Probability: the Real Thing*, on the web at //www.princeton.edu/~bayesway/, to be published by Cambridge University Press. This book, completed the day before Dick died, can be seen as his philosophical legacy, the point of arrival of the long intellectual journey he embarked on while studying under Rudolf Carnap in Chicago in the years 1946-51.

An interest in formal logic on the one hand, and probability on the other, accompanied Jeffrey throughout his career and led him to the publication of a number of influential books, including *Formal Logic: its Scope and Limits* (1967) and *Computability and Logic* (1974, with G. Boolos), in the field of formal logic; *The Logic of Decision* (1965, 1983) and *Probability and the Art of Judgment* (1992), in the field of probability, in addition to various valuable editions, inter alia the two volumes of *Studies in Inductive Logic and Probability* (volume I, edited with Carnap, 1971; volume II, 1980), which have become milestones in the literature on the foundations of probability, the edition of George Boolos' unpublished manuscripts under the title *Logic, Logic, and Logic* (1998), and the anthology of Hempel's *Selected Philosophical Essays* (2000), where some of Hempel's early articles appear in English for the first time.

Probability, and more specifically, subjective probability, was the topic on Dick's mind until his last day. Strongly convinced that "probability is a mode of judgment", he made an attempt to develop a theory of judgmental probability strong enough to satisfy the demands of rigour imposed by formal logic, but flexible enough to serve as a real guide in life, applicable to practical situations of everyday life as well as to scientific theories and hypotheses. Inspiration for this project was drawn from Bruno de Finetti's "radical probabilism" – as Dick used to call it – combining the powerful statistical tool of exchangeability with a deeply pragmatical philosophy. The resulting position that Dick defended, and developed up to his last book, maintains that "modes of judgments (in particular, probabilizing) and their attendant standards of rationality are cultural artifacts, bodies of practice modified by discovery or invention of broad features seen as grounds to stand on" (*Probability and the Art of Judgment*, Cambridge U.P., 1992, p. 11). Dick's all-round probabilism goes hand in hand with a "soft-core empiricism" – as he calls it in his last book *Subjective Probability: the Real Thing* – which admits of uncertain and incomplete evidence as a basis for condi-

tioning. This point of arrival of Jeffrey's philosophical speculation seems miles away from its beginnings with Carnap, but Dick always regarded his own work as a natural offspring of his master's teaching, and as "a further step in the development of logical empiricism, i.e., the movement, not its particular verificationist stage at ca. 1929" (*Probability and the Art of Judgment*, p. 13).

Indeed, Dick's work testifies that philosophy of science can still be done rigorously and constructively in an empiricist fashion, and that the message of logical empiricists is far from dead. To stay alive, it needs to be re-shaped along pragmatical lines and backed up by strong conceptual tools. As it was in the beginning, when the movement started, this kind of epistemology can only be performed at the price of hard work, for the backbone of "soft-core empiricism" is hard-core methodology. Dick firmly believed in this kind of programme, and his radical empiricism will remain a source of inspiration for scholars and students in the philosophy of science.

A lot more could be said about Jeffrey the philosopher, but let me add a few words about the man. Dick was a gentle and friendly kind of person; often cheerful, he liked to joke about life and people, starting with himself. He enjoyed inventing nicknames for his friends and relatives, and occasionally referred to himself as "Diamond Jim". He was fond of using witty expressions and his humour was irresistible; a conversation with him was always very rewarding, a pleasure and an enrichment. His open-mindedness and curiosity about other people's ideas was matched only by his generosity and willingness to share his own views.

Good-bye Diamond Jim, your yellow bike is no longer in the rack outside 1879 Hall, and you are no longer in your study, with the pictures of Frank Ramsey and Janina Hosiasson-Lindenbaum standing next to those of your beloved teacher Carnap and dear colleague and friend "Peter" Hempel. But your spirit, floating in the "Big Bayesian Pie", still inspires your "brothers in Bayes". That very special blend of generosity and frankness, sweetness and irony, lucidity and modesty that you represented to all your friends will not be forgotten. The "comrades and fellow travellers in the struggle for Bayesianism", as you addressed your readers in your last book, will keep your memory and your ideas alive.

Maria Carla Galavotti

INDEX OF NAMES

Not included are: Figures, Tables, Notes, References

Abraham, R. 264
Abramson, B. 208
Adams, J.B. 201
Adorno, T.W. 7
Aizerman, M.A. 238, 240
Albert, H. 7
Alchourrón, C. 229, 237
Archimedes 152
Aristotle 1f., 57, 59, 62, 65, 187, 263, 275, 278, 283
Arnol'd, V.I. 156, 263-265, 268
Arntz, G. 299
Avenarius, R. 2
Awodey, S. 332
Ayer, A.J. 334
Bacon, F. 1, 131
Bacon, R. 1
Baltas, A. V
Barlow, R.E. 255
Baron, S. 309
Bass, R. 156f.
Batens, D. 100, 102, 109
Bayes, T. 25, 53f., 57, 62, 68-76, 79, 83-90, 136f., 139, 161, 163, 199-214, 217-222, 255, 268, 354
Beller, S. 342
Bernoulli, J. 2, 145f.
Besso, M. 154
Bode, J. 141f., 155-157
Bogen, J. 252
Bohr, N. 27, 96, 126, 219
Boltzmann, L. 333, 340f.
Bolzano, B. 159, 162, 339-341
Boole, G. 74, 142
Boolos, G. 153, 173-177, 196, 353
Boutilier, C. 232, 236
Bouveresse, J. 340f.
Bovens, L. 219
Boyd Orr, J. 311, 313
Brentano, F. 332, 340f.
Bricmont, J. 6, 9, 11
Bridgman, P.W. 9
Broome, J. 228
Brown, E.W. 156
Buchanan, B.G. 199-201
Burali-Forti, C. 172
Burri, M. 342
Cantor, G. 175, 177, 179, 341
Carnap, R. 3-5, 10, 18, 62, 146, 151, 158f., 161-163, 208, 225, 265f., 269, 297, 331-339, 342, 345, 353f.
Carnot 95-97, 122, 125-127
Carrier, M. 334
Cartan, H. 197
Cartwright, N. 38, 41, 144
Carus, A.W. 332
Cassirer, E. 331, 333
Cavaillès, J. 189
Chapel, E. 314
Church, A. 334
Clark, P. 185f., 195f.
Clark, R.W. 155
Clausius 95, 97-101, 122f., 125-127
Cohen, L.J. 258
Cohen, R.S. 300f., 303, 313, 350
Cometti, J.-P. 339, 342
Comte, A. 64
Corfield, D. 208
Couturat, L. 341
Da Costa, N.C.A. 101
Darwiche, A. 236
Darwin, C. 131
Davenport, F.C. 154
Davidson, D. 311
Dawson, J. 346
Dedekind, R. 170-172, 174, 179, 188, 195
Dempster, A.P. 218
Desanti, J.-T. 189
Descartes, R. 1, 9f., 162, 186f.
Dewey, J. 2, 10
Diamond, C. 333
Dilthey, W. 331
Dingler, H. 345f.
Dokic, J. 340
Dombal, F.T. de 201, 210-212
Duda, R.O. 201
Duhem, P. 6, 9, 158
Dummett, M. 178, 196
Duns Scotus 1
Dyson, F. 154
Eddington, A.S. 154f.
Edelmann, M. 302
Edgeworth, F.Y. 227
Ehrenfels, C. von 340

Eilenberg, S. 197
Einstein, A. 9, 66, 124, 126, 153-155, 159, 187f., 264, 335
Eintalu, J. 296
Euclid 2
Euler, L. 265, 341
Exner, S. 333, 350
Faludi, A. 300, 314
Fechner, G.T. 334
Feferman, S. 343
Feigenbaum, E.A. 199
Feigl, H. 2f., 333
Festa, R. 70f.
Fetzer, J.H. 62
Feyerabend, P. 6f., 293, 351
Finetti, B. de 145-147, 152, 158, 161, 214, 222, 334, 353
Fisher, R.A. 161
Foucault, M. 341
Fraassen, B. van 37, 74-76, 78f., 84, 125
Fraenkel, A. 176
Frank, J. 310
Frank, P. 7, 9f., 156, 332f.
Franklin, A. 257
Frege, G. 9, 63, 151, 162, 169-181, 185-189, 195-197, 335, 339
Fresnel, A.J. 257
Friedman, M. 331, 333-335
Fuller, S. 10
Gabriel, G. 331
Gadol, E.T. 350
Galavotti, C.M. V, 334
Galileo, G. 1, 74
Gärdenfors, P. 135, 137, 229f., 237, 248f.
Gaschnig, J. 201
Gentzen, G. 63, 343-346
Giere, R. 28
Gillies, D.A. V, 208, 212, 217-222, 263
Gillies, D.F. 212
Gluck, M. 342
Glymour, C. 334
Gödel, K. 153f., 159, 161, 187f., 259, 332, 345f.
Gold, J.R. 301
Gonshor, H. 196
Goodman, N. 136
Granger, G.-G. 189, 341
Grattan-Guinness, I. 253, 256, 258, 263, 266f.
Grobler, A. 30, 48f., 54f.
Großmann, M. 154
Habermas, J. 7
Hacohen, M. 343
Haesaert, L. 109
Hahn, H. 332
Hajek, P. 79
Hale, B. 171f.
Halpern, J.Y. 137
Hamilton, W.R. 265, 270
Hannequin, A. 341
Hanson, N.R. 65, 127
Hansson, B. 248
Harman, G. 61f., 72
Hartmann, S. 219
Hausman, D.M. 227
Hawthorne, J. 220f.
Hayek, F.A. von 7f., 287, 340
Hayes, W. 157
Heckerman, D. 201, 218
Hegel, G.W.F. 283
Heidegger, M. 331, 336
Heidelberger, M. 333
Heisenberg, W. 126
Hempel, C.G. 2, 59, 62, 65, 69, 71f., 162, 332-334, 337, 353f.
Hermann, F.B. 282
Herschel, J. 3, 96, 123f.
Hertz, H. 340
Hesse, M. 334
Higgs 66f.
Hilbert, D. 63, 186-188, 345
Hildebrand 287
Hills, J.G. 156f.
Hintikka, J. 63, 65f., 71, 116, 152
Hodes 176
Höfler, A. 332
Hogben, L. 311-313
Holt, J. 11
Holton, G. 153
Horgan, J. 10
Horkheimer, M. 7
Hosiasson-Lindenbaum, J. 354
Howard, R.A. 205
Howson, C. 72, 89, 208
Hufeland, G. 281f., 285, 293
Hume, D. 2, 18, 136, 159-161, 169-176, 178-181, 195-197, 276, 295
Husserl, E. 159, 331f., 340f.
Irzik, G. 334
Janik, A. 342
Jaśkowski, S. 101
Jeffrey, R. VI, 162, 353f.
Jeffreys, H. 252
Jensen, R. 343
Jevons 69, 287
Joule, J.P. 125-127
Judson, P.M. 342
Julius Caesar 180
Kant, I. 3, 57, 63, 151, 160, 186-188, 190, 265, 331-333, 336, 340f.
Kaplan 334
Kaufmann, F. 273
Kelsen, H. 7
Kelvin 97

Kenny, B. 341
Kepler, J. 294
Kim, J.H. 203, 206, 333
Klein, F. 345
Knies 287
Knight, E. 313
Koertge, N. 9
Köhler, E. 141f., 152, 159
Kolmogorov, A.N. 156f., 201, 214, 263, 265
Konrad, M. 340
Kraus 27
Kräutler, H. 300
Kronecker, L. 197
Kuhn, T.S. 6, 8, 10, 124, 144, 158, 253f., 263, 334
Kuipers, T. V, 49, 53-55, 66, 71, 152
Kvasz, L. 263, 268
Ladyman, J. 78
Lagrange, J.L. 265, 270, 341
Lakatos, I. 54, 123f., 161, 191
Lamprecht, K. 7
Lanfredini, R. 331f.
Lange, F.A. 334
Lange, M. 135, 143
Langley, P. 65f.
Laplace, P.S. 69, 156
Laskar, J. 157
Laudan, L. 6, 77f., 125
Lauritzen, S.L. 203, 206f., 209f.
Leaper, D.J. 201, 211
Leibniz, G.W. 162
Lenat, D. 159
Lenin, W.I. 7f.
Levi, I. 62, 73, 248
Lewis, D.K. 39, 135
Lindley, D.V. 214
Lindström, S. 232
Liouville 265
Lipton, P. 70f., 73
Litzenberger 27
Löb 153
Lorentz, H.A. 154
Luce, R.D. 240
Mach, E. 2, 7-10, 156, 265, 333, 335, 340f.
Makinson, D. 229, 237, 249
Mandelbrot, B. 49
Mannheim, K. 7
Marsden, J. 264
Marshall, A. 278, 287
Marx, K. 281, 293, 337
Matheson, J.E. 205
Matthews, M.R. 259
Maxwell, A.N. 296
May, E. 345f.
Mayhall, C.W. 335-339
Medawar, P.B. 258
Meheus, J. 121-129, 131
Meinong, A. von 332, 340
Menger, C. 7f., 161, 278, 285-287, 293, 295, 340
Menzler-Trott, E. 343-346
Michelson, A.A. 153f., 269
Milford, K. 21, 293-295
Mill, J.S. 1f., 62, 131, 160, 278, 287
Miller 27
Miller, D. 208, 287, 296f.
Minkowski, H. 154
Mises, R. von 209, 287, 333
Modigliani 27
Monton, B. 71, 79
Morley, D.W. 303, 309-311
Morley, E. 153f., 270
Morris, C. 10
Moser, J.K. 156, 263, 265
Moulin, H. 238
Mulligan, K. 339-342
Murray, C. 155
Musil, R. 339f.
Napier, J. 155
Napoleon Bonaparte 59, 68
Nayak, A.C. 137
Neapolitan, R.E. 203, 210
Neumann, J. von 176
Neurath, Otto 6-8, 10, 18, 158, 225-227, 229, 237, 241, 299-315, 332, 336f., 339
Newcomb, S. 156
Newell, A. 159
Newton, I. 78, 123f., 131, 154, 156f., 160, 251-257, 263-269, 275, 335
Neyman, J. 62, 161, 163, 209
Ng, K. 208
Nieto, M.M. 155
Nietzsche, F. 163
Niiniluoto, I. V, 28, 59, 62, 64, 66, 68, 70f., 73-78, 83-85, 89, 136, 152, 158
Noether, E. 265
Ockham, W. of 1
Ogden, C.K. 312
Okasha, S. 70f.
Oppenheim, P. 2
Ovenden, M.W. 156f.
Pais, A. 153-155, 159
Panofsky, W. 49
Parrini, P. 331, 334f.
Paty, M. V
Peano, G. 169-171, 195
Pearl, J. 203-207, 209, 219, 236
Pearson, E.S. 62, 161, 163
Peirce, C.S. 28, 33, 57-66, 68f., 72-75, 78, 127
Pfabigan, A. 343
Phillips, M. 49
Planck, M. 333

Plato 161, 171, 265, 332
Plato, J. von 343
Poincaré, H. 9, 156, 178, 186-188, 265, 335f., 340f.
Pollock, J. 90
Popper, K.R. 2-4, 6-8, 11, 18-22, 31, 62, 64, 70, 129, 144, 152, 160-163, 191, 208, 212-214, 217f., 220-222, 251-259, 263, 265f., 268-270, 273f., 287, 293, 295-297, 340, 343
Potter, M. 180
Poyla, G. 159
Priest, G. 101
Psillos, S. 59, 78
Quine, W.V.O. 4, 6, 9, 126, 235, 339
Raiffa, H. 240
Ramsey, F. 143, 161, 222, 247f., 334, 348, 354
Ranke, L. 283
Rau, K.H. 282
Ravetz, J.R. 259
Rédei, M. V
Reichenbach, H. 3-5, 18, 61f., 191, 206f., 296, 333-335
Reidemeister-Neurath, M. 299-301, 303-305, 310-313
Reilly, C.H. 302, 306, 308, 310
Reisch, G. 332
Renner 342
Renouvier 341
Requard, F. 345
Rescher, N. 254
Ricardo, D. 280
Richtmyer, F.K. 256
Riehl, A. 334
Riemann, B. 154, 186, 270
Roscher, W. 278, 281-287, 293f.
Rosenberg, A. 227f.
Rosenthal-Schneider, I. 155
Rotha, P. 311, 313f.
Rott, H. 232-234, 236, 238, 247-249
Roy, A.E. 156f.
Rusnock, P. 341
Russell, B. 63, 143, 160, 169-171, 175f., 178, 181, 196f., 335, 339f.
Rutherford 27
Ryckman, T. 333, 335
Sagan, C. 157f.
Salmon, M.H. 331
Salmon, W.C. 2, 28, 62, 65, 70, 73, 331, 334f.
Sarmany-Parsons 343
Sartre, J.P. 341
Savage 161, 209, 214
Scheck, F. 157
Scheler, M. 7
Schilpp, P.A. 161
Schlick, M. 225f., 229, 241, 331-337, 342, 348f.
Schmitt, C. 7
Schmoller, G. 7f., 286f.
Schorske, C.E. 342f.
Schumpeter, J. 287, 342
Schuppe, W. 2
Schurz, G. 65f.
Scott, T. 157
Sen, A. 240
Shafer, G. 218
Shapley, H. 156-158
Shedel, J. 342
Shewhart 255
Shimony 161
Shortliffe, E.H. 199-201
Siegel, C.L. 156
Simon, H. 151, 159, 227
Simons, P. 340
Smith, A. 281f., 293
Smokler, H. 68
Sokal, A. 6, 9, 11
Sommerfeld, A. 27
Spector, S. 343
Spiegelhalter, D.J. 203, 206f., 209f.
Spinoza, B. 189
Spirtes, P. 221
Spohn, W. 135-137, 140, 151-163, 227
Stadler, F. V, 17f., 22
Steck, M. 345f.
Stegmüller, W. 4
Stern, D. 333
Stöltzner, M. 333
Sucar, L.E. 212
Sussman, G.J. 157
Szabo, M. 343
Tarski, A. 234, 256-259, 337, 339f., 344
Taylor, S. 311
Thagard, P. 28
Titius, J.D. 155
Tuomela, R. 68, 70f., 158
Uebel, T.E. 332f.
Ullian, J.S. 235
Urbach, P. 72, 208
Van der Waals 27
Velikovsky, I. 157
Venn, J. 60
Vihan, P. 346
Waismann, F. 337
Ward, S.V. 301f.
Weber, M. 7
Weinberg, S. 8, 10
Weyl, H. 11, 187, 345
Whewell, W. 3, 64, 73, 124, 131, 295
Whitehead, A.N. 335
Williams, A.V. 300f., 303, 305, 310f.
Williamson, J. 208, 219

Wilson, F. 160f.
Wisdom, J. 156f.
Wiśniewski, A. 49, 54f.
Wittgenstein, L. 159, 333, 335-337, 339-341
Wolenski, J. 340
Wolfe, L. 302
Wolters, G. V, 332, 345
Woodward, W. 252
Worrall, J. 253
Wright, C. 171f., 174f., 177-179, 196f.
Zahar, E. 191
Zermelo, E. 169, 176, 197
Ziman, J. 255
Zimmermann, R. 341

VIENNA CIRCLE INSTITUTE YEARBOOK

The *Vienna Circle Institute* is devoted to the critical advancement of science and philosophy in the broad tradition of the Vienna Circle, as well as to the focussing of cross-disciplinary interest on the history and philosophy of science. The Institute's *Yearbooks* provide a forum for the discussion of exact philosophy, logical and empirical investigations, and analysis of language. Each volume centers around a special topic which is complemented with a permanent section with essays arising from the scientific activities at the Institute and reviews of recent works in the history of philosophy of science or others with a particular relation to the tradition of logical empiricism.

1 [1993] F. STADLER (ed.), *Scientific Philosophy: Origins and Developments*. 1993.
ISBN 0-7923-2526-5

2 [1994] H. PAUER-STUDER (ed.), *Norms, Values, and Society*. 1994.
ISBN 0-7923-3071-4

3 [1995] W. DEPAULI-SCHIMANOVICH / E. KÖHLER / F. STADLER (eds.), *The Foundational Debate. Complexity and Constructivity in Mathematics and Physics*. 1995.
ISBN 0-7923-3737-9

4 [1996] E. NEMETH / F. STADLER (eds.), *Encylopedia and Utopia. The Life and Work of Otto Neurath (1882-1945)*. 1996.
ISBN 0-7923-4161-9

5 [1997] W. LEINFELLNER / E. KÖHLER (eds.), *Game Theory, Experience, Rationality. Foundations of Social Sciences, Economics and Ethics. In honor of John C. Harsanyi*. 1998.
ISBN 0-7923-4943-1

6 [1998] J. WOLEŃSKI / E. KÖHLER (eds.), *Alfred Tarski and the Vienna Circle. Austro-Polish Connections in Logical Empiricism*. 1999.
ISBN 0-7923-5538-5

7 [1999] D. GREENBERGER / W.L. REITER / A. ZEILINGER (eds.), *Epistemological and Experimental Perspectives on Quantum Physics*. 1999.
ISBN 0-7923-6338-X

8 [2000] M. RÉDEI / M. STÖLTZNER (eds.), *John von Neumann and the Foundations of Quantum Physics*. 2001.
ISBN 0-7923-6812-6

9 [2001] M. HEIDELBERGER / F. STADLER (eds.), *History of Philosophy of Science. New Trends and Perspectives*. 2002.
ISBN 1-4020-0509-1

10 [2002] F. STADLER (ed.), *The Vienna Circle and Logical Empiricism*. 2003.
ISBN 1-4020-1269-1

11 [2003] F. STADLER (ed.), *Induction and Deduction in the Sciences*. 2004.

Springer Science+Business Media, B.V.