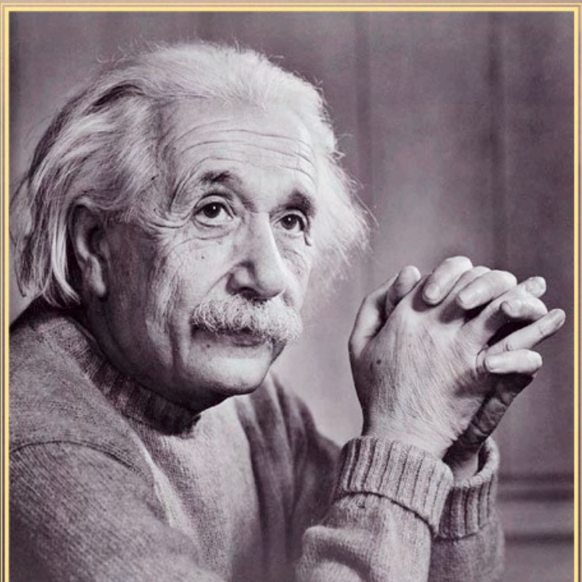
Handbook of
THE PHILOSOPHY OF SCIENCE

General Editors: DOV M. GABBAY, PAUL THAGARD, AND JOHN WOODS

PHILOSOPHY *of* PHYSICS

PART B



Edited by Jeremy Butterfield
and John Earman

N·H

Philosophy of Physics
Part B

Handbook of the Philosophy of Science

General Editors

Dov M. Gabbay
Paul Thagard
John Woods

Cover: Photograph of Albert Einstein taken in 1948 at Princeton
by Yousuf Karsh

ELSEVIER

AMSTERDAM • BOSTON • HEIDELBERG • LONDON • NEW YORK • OXFORD
PARIS • SAN DIEGO • SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO
North-Holland is an imprint of Elsevier

N·H

Philosophy of Physics
Part B

Edited by

Jeremy Butterfield
All Souls College, University of Oxford, Oxford, UK

and

John Earman
Department of History and Philosophy of Science,
University of Pittsburgh, Pittsburgh, PA, USA

ELSEVIER
AMSTERDAM • BOSTON • HEIDELBERG • LONDON • NEW YORK • OXFORD
PARIS • SAN DIEGO • SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO
North-Holland is an imprint of Elsevier
N·H

North-Holland is an imprint of Elsevier
Radarweg 29, PO Box 211, 1000 AE Amsterdam, The Netherlands
The Boulevard, Langford Lane, Kidlington, Oxford OX5 1GB, UK

First edition 2007

Copyright © 2007 Elsevier B.V. All rights reserved

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means electronic, mechanical, photocopying, recording or otherwise without the prior written permission of the publisher

Permissions may be sought directly from Elsevier's Science & Technology Rights Department in Oxford, UK: phone (+44) (0) 1865 843830; fax (+44) (0) 1865 853333; email: permissions@elsevier.com. Alternatively you can submit your request online by visiting the Elsevier web site at http://elsevier.com/locate/permissions, and selecting *Obtaining permission to use Elsevier material*

Notice
No responsibility is assumed by the publisher for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions or ideas contained in the material herein. Because of rapid advances in the medical sciences, in particular, independent verification of diagnoses and drug dosages should be made

Library of Congress Cataloging-in-Publication Data
A catalog record for this book is available from the Library of Congress

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

ISBN-13 (Set): 978-0-444-51560-5
ISBN-10 (Set): 0-444-51560-7
ISBN-13 (Part A): 978-0-444-53001-1
ISBN-10 (Part A): 0-444-53001-0
ISBN-13 (Part B): 978-0-444-53002-8
ISBN-10 (Part B): 0-444-53002-9

For information on all North-Holland publications
visit our website at books.elsevier.com

Printed and bound in The Netherlands

07 08 09 10 11 10 9 8 7 6 5 4 3 2 1

GENERAL PREFACE

Dov Gabbay, Paul Thagard, and John Woods

Whenever science operates at the cutting edge of what is known, it invariably runs into philosophical issues about the nature of knowledge and reality. Scientific controversies raise such questions as the relation of theory and experiment, the nature of explanation, and the extent to which science can approximate to the truth. Within particular sciences, special concerns arise about what exists and how it can be known, for example in physics about the nature of space and time, and in psychology about the nature of consciousness. Hence the philosophy of science is an essential part of the scientific investigation of the world.

In recent decades, philosophy of science has become an increasingly central part of philosophy in general. Although there are still philosophers who think that theories of knowledge and reality can be developed by pure reflection, much current philosophical work finds it necessary and valuable to take into account relevant scientific findings. For example, the philosophy of mind is now closely tied to empirical psychology, and political theory often intersects with economics. Thus philosophy of science provides a valuable bridge between philosophical and scientific inquiry.

More and more, the philosophy of science concerns itself not just with general issues about the nature and validity of science, but especially with particular issues that arise in specific sciences. Accordingly, we have organized this Handbook into many volumes reflecting the full range of current research in the philosophy of science. We invited volume editors who are fully involved in the specific sciences, and are delighted that they have solicited contributions by scientifically-informed philosophers and (in a few cases) philosophically-informed scientists. The result is the most comprehensive review ever provided of the philosophy of science.

Here are the volumes in the Handbook:

Philosophy of Science: Focal Issues, edited by Theo Kuipers.

Philosophy of Physics, edited by Jeremy Butterfield and John Earman.

Philosophy of Biology, edited by Mohan Matthen and Christopher Stephens.

Philosophy of Mathematics, edited by Andrew Irvine.

Philosophy of Logic, edited by Dale Jacquette.

Philosophy of Chemistry and Pharmacology, edited by Andrea Woody and Robin Hendry.

Philosophy of Statistics, edited by Prasanta S. Bandyopadhyay and Malcolm Forster.

Philosophy of Information, edited by Pieter Adriaans and Johan van Benthem.

Philosophy of Technological Sciences, edited by Anthonie Meijers.

Philosophy of Complex Systems, edited by Cliff Hooker and John Collier.

Philosophy of Earth Systems Science, edited by Bryson Brown and Kent Peacock.

Philosophy of Psychology and Cognitive Science, edited by Paul Thagard.

Philosophy of Economics, edited by Uskali Mäki.

Philosophy of Linguistics, edited by Martin Stokhof and Jeroen Groenendijk.

Philosophy of Anthropology and Sociology, edited by Stephen Turner and Mark Risjord.

Philosophy of Medicine, edited by Fred Gifford.

Details about the contents and publishing schedule of the volumes can be found at http://www.johnwoods.ca/HPS/.

As general editors, we are extremely grateful to the volume editors for arranging such a distinguished array of contributors and for managing their contributions. Production of these volumes has been a huge enterprise, and our warmest thanks go to Jane Spurr and Carol Woods for putting them together. Thanks also to Andy Deelen and Arjen Sevenster at Elsevier for their support and direction.

ACKNOWLEDGEMENTS

It is a pleasure to thank several people. First and foremost, we are grateful to John Woods, Dawn Collins, Jane Spurr and their colleagues at Lethbridge, Alberta, and London, England for their unflagging support and encouragement. We also thank Brian Hepburn for invaluable help with references and the preparation of the text.

This page intentionally left blank

CONTENTS

Part B

This volume is dedicated to the memories of

Robert Clifton (1964–2002)

and

James Cushing (1937–2002).

This page intentionally left blank

INTRODUCTION

Jeremy Butterfield and John Earman

1 THE PHILOSOPHY OF PHYSICS TODAY

In the last forty years, philosophy of physics has become a large and vigorous branch of philosophy, and so has amply won its place in a series of Handbooks in the philosophy of science. The reasons for its vigour are not far to seek. As we see matters, there are two main reasons; the first relates to the formative years of analytic philosophy of science, and the second to the last forty years.

First, physics had an enormous influence on the early phase of the analytic movement in philosophy. This influence does not just reflect the fact that for the logical positivists and logical empiricists, and for others such as Russell, physics represented a paradigm of empirical knowledge. There are also much more specific influences. Each of the three main pillars of modern physics — thermal physics, quantum theory and relativity — contributed specific ideas and arguments to philosophical debate. Among the more obvious influences are the following.

Thermal physics and the scientific controversy about the existence of atoms bore upon the philosophical debate between realism and instrumentalism; and the rise of statistical mechanics fuelled the philosophy of probability. As to quantum theory, its most pervasive influence in philosophy has undoubtedly been to make philosophers accept that a fundamental physical theory could be indeterministic. But this influence is questionable since, as every philosopher of science knows (or should know!), indeterminism only enters at the most controversial point of quantum theory: viz., the alleged "collapse of the wave packet". In any case, the obscurity of the interpretation of quantum theory threw not only philosophers, but also the giants of physics, such as Einstein and Bohr, into vigorous debate: and not only about determinism, but also about other philosophical fundamentals, such as the nature of objectivity. Finally, relativity theory, both special and general, revolutionized the philosophy of space and time, in particular by threatening neo-Kantian doctrines about the nature of geometry.

These influences meant that when the analytic movement became dominant in anglophone philosophy, the interpretation of modern physics was established as a prominent theme in its sub-discipline, philosophy of science. Accordingly, as philosophy has grown, so has the philosophy of physics.

But from the 1960s onwards, philosophy of physics has also grown for a reason external to philosophy. Namely, within physics itself there has been considerable

Handbook of the Philosophy of Science. Philosophy of Physics
Volume editors: Jeremy Butterfield and John Earman
General editors: Dov M. Gabbay, Paul Thagard and John Woods
© 2007 Elsevier B.V. All rights reserved.

interest in foundational issues, with results that have many suggestive repercussions for philosophy. Again, there have been various developments within physics, and thereby various influences on philosophy. The result, we believe, is that nowadays foundational issues in the fundamental physical theories provide the most interesting and important problems in the philosophy of physics. We have chosen the topics for this volume in accord with this conviction. In the next Subsection, we will articulate some of these foundational issues, and thereby introduce the Chapters of the volume.

2 CURRENT FOUNDATIONAL ISSUES IN PHYSICS

We will first discuss these issues under five headings. The first three correspond to the three pillars of modern physics mentioned in Section 2.1; i.e. thermal physics, quantum theory and relativity theory. The fourth and fifth concern combinations of these pillars; and lead to speculations about the future of physics. These five headings will provide a way of introducing most of this volume's Chapters, albeit not in the order in which they occur. Then, after these five headings, we will introduce the volume's remaining two Chapters.

2.1 Thermal physics

Controversies about the foundations of thermal physics, especially the characterization of the approach to equilibrium, have continued unabated since the days of the field's founding fathers, such as Maxwell and Boltzmann. Some aspects of the original controversies can be seen again in modern discussions. But the controversies have also been transformed by the development of several scientific fields; especially the following three, which have grown enormously since the 1960s:

(i) classical mechanics, and its offspring such as ergodic theory and chaos theory;

(ii) quantum thermal physics; and

(iii) cosmology, which nowadays provides a very detailed and so fruitful context for developing and evaluating Boltzmann's bold idea that the ultimate origin of the "arrow of time" is cosmological.

In this volume, the foundations of thermal physics is represented by the Chapters by Uffink and by Emch, who cover classical and quantum aspects, respectively. Among the topics Uffink discusses, two receive special attention: the evolution of Boltzmann's views, and the mathematical framework of stochastic dynamics. Emch adopts the formalism of algebraic quantum statistical mechanics, and reviews many results about that formalism's notion of equilibrium, i.e. KMS states. Two other Chapters also provide a little stage-setting for Uffink and Emch, though without pursuing the relation to thermal physics: viz. the Chapters by Butterfield on classical mechanics, and by Ellis on cosmology.

2.2 Quantum theory

Since the 1960s, the physics community has witnessed a revival of the debates about the interpretation of quantum theory that raged among the theory's founding fathers. In the general physics community, the single most influential author has no doubt been John Bell, not only through his non-locality theorem and the many experiments it engendered, but also through his critique of the "Copenhagen orthodoxy" and his sympathy towards the pilot-wave and dynamical collapse heterodoxies. But in more specialist communities, there have been other crucial factors that have animated the debate. Mathematical physicists have developed a deep understanding of the various relations between quantum and classical theories. Since the 1970s, there has been progress in understanding decoherence, so that nowadays, almost all would accept that it plays a crucial role in the emergence of the classical world from quantum theory. And since the 1990s, the burgeoning fields of quantum information and computation have grown out of the interpretative debates, especially the analysis of quantum non-locality.

In this volume, these topics are taken up by Dickson, Landsman and Bub. Dickson surveys the formalism of non-relativistic quantum theory, and some of the main interpretative issues, including empirical content, quantum uncertainty, the measurement problem, and non-locality. For the most part, Landsman reviews from the perspective of mathematical physics the relations between quantum and classical theories. In particular, he discusses various approaches to quantization and the rigorous treatments of the classical limits $\hbar \to 0$ and $N \to \infty$. But Landsman also includes discussions of the Copenhagen interpretation and decoherence. Finally, Bub presents some central ideas and results about quantum information and quantum computation. As a backdrop to this, he also briefly reviews classical information and computation; and he ends by proposing some provocative morals about the interpretation of quantum theory.

2.3 Relativity theory

The decades since the 1960s have seen spectacular developments, for both theory and experiment, in general relativity and cosmology. But this Renaissance has also been very fruitful as regards foundational and philosophical issues. Mathematical relativists have continued to deepen our understanding of the foundations of general relativity: foundations which, as mentioned in Section 1, were recognized already in the 1920s as crucial for the philosophy of space and time. And the recent transformation of cosmology from a largely speculative enterprise into a genuine science has both brought various philosophical questions closer to scientific resolution, and made other philosophical questions, e.g. about method and explanation in cosmology, much more pressing.

In this volume, these topics are represented by the Chapters by Malament, Belot and Ellis. Malament first expounds classical relativity. Then he discusses three special topics: the definition of simultaneity in special relativity, the geometrization of Newtonian gravity, and the extent to which causal structure determines spacetime geometry. Belot's main aim is to give a clear statement of the "problem of time" as it occurs in classical general relativity; and to do that, he first reviews the way time is represented in simpler classical theories, including mechanics. (Belot's Chapter thereby complements Butterfield's: both expound aspects of classical Hamiltonian theories, and stress how some of these aspects reappear in quantum theories.) Ellis first reviews the present state of relativistic cosmological theory and its observational basis; and then investigates nine philosophical themes, including the anthropic principle and the possible existence of multiverses.

So much by way of introducing some foundational issues, and this volume's corresponding Chapters, arising *within* one of the three pillars: thermal physics, quantum theory and relativity. We turn to issues arising from combining the pillars — or rather, parts of them! We have already adumbrated the combination of the first and second: viz., in quantum thermal physics, reviewed here by Emch. It is the combination of the second and third — quantum theory and relativity — which we must now address. We shall do so under two headings, corresponding to the distinction between special and general relativity. The first corresponds, of course, to quantum field theory, which forms such a deep and well-established framework for particle physics. The second corresponds to the quantum theory of gravity — which unfortunately still remains only a hope and a goal.[1]

2.4 Quantum field theory

Although there are relativistic quantum mechanical theories of a fixed number of particles, by far the most important framework combining quantum theory and special relativity is quantum field theory. Broadly speaking, the foundational issues raised by quantum field theory differ from quantum theory's traditional interpretative issues, about measurement and non-locality (cf. *Quantum theory*, §2.2 above). There are two points here.

(i) Although quantum field theory of course illustrates the latter issues just as much as elementary quantum theory does, it apparently cannot offer a resolution of them. The measurement problem and the puzzles about non-locality arise so directly from the unitarity and tensor-product features of quantum theories, as to be unaffected by the extra mathematical structures

[1] Our image of three pillars prompts the question: what about the combination of thermal physics and relativity? When Einstein's special theory of relativity won acceptance, the rush was on to revise the various branches of classical physics to make them properly relativistic. In the case of thermodynamics, this program produced disputes about the Lorentz transformation properties of the thermodynamic quantities of heat, temperature and entropy that persisted well into the 1970s; (see [Liu, 1994] for an overview of this debate). As for classical general relativity theory, there does not currently exist a statistical mechanics that incorporates the "gravitational entropy of the universe", and it seems unlikely that there can be such a theory. But for all anyone knows, the ideas of thermal physics may play a crucial role in the hoped-for quantum theory of gravity. There are hints to that effect from, for example, black hole thermodynamics, the Unruh effect, and Hawking radiation. These topics are discussed briefly in Rovelli's chapter.

and physical ideas supplied by quantum field theory.[2] And accordingly, it has seemed to most workers to be wisest to pursue the traditional interpretative issues within non-relativistic quantum theory: if you identify a problem in a simple context, but are confident that it is not an artefact of the context's simplicity, it is surely wisest to attack it there. (And as shown in this volume by Dickson's and Landsman's Chapters, that context is by no means "really simple": non-relativistic quantum theory, and its relation to classical theories, provides an abundance of intricate structure to investigate.)

(ii) On the other hand, there are several foundational issues that are distinctive of quantum field theory. Perhaps the most obvious ones are: the nature of particles (including the topic of localization), the interpretation of renormalization, the interpretation of gauge structure, and the existence of unitarily equivalent representations of the canonical commutation relations.

In this volume, these topics are taken up by 't Hooft and by Halvorson and Müger. First, 't Hooft provides an authoritative survey of quantum field theory, from the perspective of particle physics. Among the main topics he expounds are: the quantization of scalar and spinor fields, Feynman path integrals, the ideas of gauge fields and the Higgs mechanism, renormalization, asymptotic freedom and confinement.

Halvorson and Müger discuss a smaller and distinctively foundational set of issues, using the apparatus of algebraic quantum field theory. (So their use of the algebraic approach complements the uses made by Emch and Landsman.) They discuss the nature of particles and localization, non-locality, the assignment of values to quantities (i.e. the measurement problem) and the definability of quantum fields at individual spacetime points. But they devote most of their effort to the Doplicher-Haag-Roberts theory of superselection. This theory yields deep insights into crucial structures of quantum field theory: in particular, the set of representations, the relation between the field and observable algebras, and gauge groups.

2.5 Quantum gravity

Finally, we turn to the combination of quantum theory with general relativity: i.e., the search for a quantum theory of gravity. Here there is of course no established theory, nor even a consensus about the best approach for constructing one. Rather there are various research programmes that often differ in their technical aims, as well as their motivations and conceptual frameworks. In this situation, various

[2] In some respects relativistic QFT makes the measurement problem worse. In non-relativistic quantum mechanics, the collapse of the state vector is supposed to happen instantaneously; so in the relativistic setting, one would have to develop some appropriate analogue. On the other hand, the modal interpretation of ordinary QM — which arguably provides the best hope for a no-collapse account of quantum measurement — faces formidable obstacles in relativistic quantum field theory; (see [Clifton, 2000] and Halvorson and Müger, this volume, Section 5).

foundational issues about the "ingredient" theories are cast in a new light. For example, might quantum gravity revoke orthodox quantum theory's unitarity, and thereby *en passant* solve the measurement problem? And does the general covariance (diffeomorphism invariance) of general relativity represent an important clue about the ultimate quantum nature of space and time?

In this volume, these and related questions are taken up by Rovelli. He also presents details about other topics: for example, the subject's history, the two main current programmes (string theory and loop quantum gravity), and quantum cosmology. Ellis' Chapter also discusses quantum cosmology. In this way, and indeed by addressing other fundamental questions about the idea of an "ultimate" physical theory, Ellis's Chapter provides a natural complement to Rovelli's.

So much by way of introducing Chapters that correspond to our initial three pillars of modern physics, or to combinations of them. We turn to introducing the volume's remaining two Chapters. Here our intention has been to provide Chapters whose discussions bridge the divisions between physical theories, and even those between our three pillars. In this connection, it seemed to us that of the various possible themes for such a cross-cutting discussion, the two most appropriate ones were determinism and symmetry.[3]

Accordingly, Earman discusses how determinism fares in a wide class of theories: his examples range from classical mechanics to proposals for quantum gravity. He also addresses the relations between determinism and other issues: in particular, predictability, the nature of spacetime, and symmetry. Symmetry in classical physics is given a wide-ranging survey by Brading and Castellani. Among other topics, they discuss: Curie's principle, the advent of group theory into physics, canonical transformation theory, general covariance in general relativity, and Noether's theorems. Various aspects of symmetry and invariance in quantum physics are discussed in the Chapters by Dickson, Emch, Halvorson, and Landsman. But a synoptic overview of this complex topic remains to be written — which we hope will be taken as a challenge by some of our readers.

Let us sum up this introduction to the Chapters that follow, with two comments that are intended to give the prospective reader — perhaps daunted by the many pages ahead! — some courage.

First, it is obvious that by our lights, there is no sharp line between philosophy of physics and physics itself. So it is no surprise that some of the best work in philosophy of physics is being done by physicists (as witnessed by several contributions to this volume). No surprise: but certainly, to be welcomed. Conversely, to the traditionally trained philosopher, work by philosophers of physics is liable to look more like physics than philosophy. But for us, this blurring of disciplinary boundaries is no cause for concern. On the contrary, it represents an opportunity for philosophy to enrich itself. And in the other direction, philosophers can hope

[3]Other good candidates include the "direction of time", or irreversibility, and the constitution of matter. But adding chapters on these or other cross-cutting themes would have made the volume altogether too long.

that the foundations, and even philosophy, of physics can be a source of heuristic ideas for physics. Or at least, physicists' interest in foundational questions now offers philosophers of physics the opportunity of fruitful discussion with physicists.

But agreed: this enrichment of philosophy does not come for free. And the need to master technical material which is often difficult can be a barrier to entering the philosophy of physics. In designing this volume, our response to this problem has of course been, not to try to lower the barrier, at the cost of scholarship and of fostering illusory hopes: rather our strategy has been to commission Chapters that cover their chosen topics as expertly and completely as possible. So to the reader, our message is simple: take heart! Once you are over the barrier, new vistas open for the philosophy of science.

3 OUTLOOK: HALFWAY THROUGH THE WOODS

Finally, we would like to set the stage for this volume, by making two connected comments about the present state of fundamental physics. Though it may seem naive or hubristic for philosophers to make such comments, we believe it is worth the risk. For we think that at the present juncture fundamental physics is unusually open to contributions from philosophical reflection; and it will be clear from our comments that together they represent an invitation to the reader to make such contributions! The first comment concerns the amazing successes of present-day physics; the second, the fact that so much remains to be understood.

3.1 Successes

First, we want to celebrate the extraordinary achievements of modern physics; specifically of quantum theory and relativity theory. We propose to do this by emphasising how contingent, indeed surprising, it is that the basic postulates of relativity and quantum theory have proved to be so successful in domains of application far beyond their originally intended ones.

Examples are legion. We pick out two examples, almost at random. Why should the new chronogeometry introduced by Einstein's special relativity in 1905 for electromagnetism, be extendible to mechanics, thermodynamics and other fields of physics? And why should the quantum theory, devised for systems of atomic dimensions (10^{-8} cm) be good both for scales much smaller (cf. the nuclear radius of ca. 10^{-12} cm) and vastly larger (cf. superconductivity and superfluidity, involving scales up to 10^{-1} cm)? Indeed, much of the history of twentieth century physics is the story of the consolidation of the relativity and quantum revolutions: the story of their basic postulates being successfully applied ever more widely.

The point applies equally well when we look beyond terrestrial physics. We have in mind, first, general relativity. It makes a wonderful story: the theory was created principally by one person, motivated by conceptual, in part genuinely philosophical, considerations — yet it has proved experimentally accurate in all kinds of astronomical situations. They range from weak gravitational fields such as occur in the solar system — here it famously explains the minuscule portion of the precession of the perihelion of Mercury (43" of arc per century) that was unaccounted for by Newtonian theory; to fields 10,000 times stronger in a distant binary pulsar — which in the last twenty years has given us compelling (albeit indirect) evidence for a phenomenon (gravitational radiation) that was predicted by general relativity and long searched for; and to exotic objects such as black holes. But general relativity is not the only case. Quantum theory has also been extraordinarily successful in application to astronomy: the obvious example is the use of nuclear physics to develop a very accurate and detailed theories of nucleosynthesis in the very early universe, and of stellar structure and evolution.

Indeed, there is a more general point here, going beyond the successes of relativity and quantum theory. Namely, we tend to get used to the various unities in nature that science reveals — and thereby to forget how contingent and surprising they are. Of course, this is not just a tendency of our own era. For example, nineteenth century physics confirmed Newton's law of gravitation to apply outside the solar system, and discovered terrestrial elements to exist in the stars (by spectroscopy): discoveries that were briefly surprising, but soon taken for granted, incorporated into the educated person's 'common sense'. Similarly nowadays: the many and varied successes of physics in the last few decades, in modelling very accurately phenomena that are (i) vastly distant in space and time, and-or (ii) very different from our usual laboratory scales (in their characteristic values of such quantities as energy, temperature, or pressure etc.), reveal an amazing unity in nature. General theoretical examples of such unity, examples that span some 200 years, are: the ubiquitous fruitfulness of the field concept; and more specifically, of least action principles. For a modern, specific (and literally spectacular) example, consider the precision and detail of our models of supernovae; as confirmed by the wonderful capacity of modern telescope technology to see and analyse individual supernovae, even in other galaxies.

3.2 Clouds on the horizon

And yet: complacency, let alone triumphalism, is not in order! Current physics is full of unfinished business — that is always true in human enquiry. But more to the point, there are clouds on the horizon that may prove as great a threat to the continued success of twentieth century physics, as were the anomalies confronting classical physics at the end of the nineteenth century.

Of course, people differ about what problems they find worrisome; and among the worrisome ones, about which problems are now ripe for being solved, or at least worth addressing. As philosophers, we are generalists: so we naturally find all the various foundational issues mentioned above worrisome. But being generalists, we will of course duck out of trying to say which are the closest to solution, or which are most likely to repay being addressed! In any case, such judgments are hard to adjudicate, since intellectual temperament, and the happenstance of what one knows about or is interested in, play a large part in forming them.

But we would like to end by returning to one of Section 2's "clouds": a cloud which clearly invites philosophical reflection, and perhaps contributions. Namely, the problem of quantum gravity; in other words, the fact that general relativity and quantum theory are yet to be reconciled. As mentioned in Section 2.5, Rovelli (this volume) discusses how the contrasting conceptual structures of the "ingredient" theories and the ongoing controversies about interpreting them, make for conflicting basic approaches to quantum gravity.

But we want here to emphasise another reason why we still lack a successful theory, despite great effort and ingenuity. In short, it is that the successes of relativity and quantum theory, celebrated in Comment 3.1 above, conspire to deprive us of the relevant experimental data.

Thus there are general reasons to expect data characteristic of quantum gravity to arise only in a regime of energies so high (correspondingly, distances and times so short) as to be completely inaccessible to us. To put the point in terms of length: the value of the Planck length which we expect to be characteristic of quantum gravity is around 10^{-33} cm. This is truly minuscule: the diameters of an atom, nucleus, proton and quark are, respectively, about 10^{-8}, 10^{-12}, 10^{-13}, and 10^{-16} cm. So the Planck length is as many orders of magnitude from the (upper limit for) the diameter of a quark, as that diameter is from our familiar scale of a centimetre!

We can now see how quantum gravity research is in a sense the victim of the successes of relativity and quantum theory. For those successes suggest that we will not see any "new physics" intimating quantum gravity even at the highest energies accessible to us. The obvious example is quasars: these are typically a few light-days in diameter, and yet have a luminosity 1000 times that of our galaxy (itself 100,000 light-years across, containing a hundred billion stars). They are the most energetic, distant (and hence past!) celestial objects that we observe: they are now believed to be fuelled by massive black holes in their cores. Yet suggestions, current thirty years ago, that their stupendous energies and other properties that we *can* observe, could only be explained by fundamentally new physics, have nowadays given way to acceptance that "conventional physics" describing events *outside* the black hole's event-horizon can do so. (Agreed, we expect the physics deep inside the black hole, in the vicinity of its singularity, to exhibit quantum gravity effects: but if ever a region deserved the name "inaccessible", this is surely one!) So the situation is ironic, and even frustrating: quantum gravity research is a victim of its ingredient theories' success.

In any case, the search for quantum gravity is wide open. In closing, we would like to endorse an analogy of Rovelli's [1997]. He suggests that our present search is like that of the mechanical philosophers such as Galileo and Kepler of the early seventeenth century. Just as they struggled with the clues given by Copernicus and Brahe, *en route* to the synthesis given by Newton, so also we are "halfway through the woods". Of course we should be wary of too grossly simplifying and periodizing the scientific revolution, and *a fortiori* of facile analogies between different historical situations. Nevertheless, it is striking what a "mixed bag" the doctrines of figures such as Galileo and Kepler turn out to have been, from the perspective of the later synthesis. For all their genius, they appear to us (endowed with the anachronistic benefits of hindsight), to have been "transitional figures". One cannot help speculating that to some future reader of twentieth century physics, enlightened by some future synthesis of general relativity and quantum theory, our current and recent efforts in quantum gravity will seem strange: worthy and sensible from the authors' perspective (one hopes), but a hodge-podge of insight and error from the reader's!

BIBLIOGRAPHY

[Clifton, 2000] R. Clifton. The modal interpretation of algebraic quantum field theory. *Physics Letters A*, *271*, 167-177, 2000.

[Liu, 1994] C. Liu. Is there a relativistic thermodynamics? A case study in the meaning of special relativity. *Studies in the History and Philosophy of Science*, *25*, 983-1004, 1994.

[Rovelli, 1997] C. Rovelli. Halfway through the woods. In J. Earman and J. Norton (Eds.), *The Cosmos of Science* (pp. 180-223). Pittsburgh: University of Pittsburgh Press and Konstanz: Universitäts Verlag, 1997.

CONTRIBUTORS

Gordon Belot
Department of Philosophy, University of Pittsburgh, 1001 Cathedral of Learning, Pittsburgh, PA 15260, USA.
gbelot@pitt.edu

Katherine Brading
Department of Philosophy, University of Notre Dame Notre Dame, 100 Malloy Hall, Indiana 46556, USA.
kbrading@nd.edu

Jeffrey Bub
Department of Philosophy, University of Maryland, College Park, MD 20742, USA.
jbub@umd.edu

Jeremy Butterfield
Trinity College, Cambridge, CB2 1TQ, UK.
jb56@cus.cam.ac.uk

Elena Castellani
Department of Philosophy, University of Florence, via Bolognese 52, 50139, Firenze, Italy.
elena.castellani@unifi.it

Michael Dickson
Department of Philosophy, University of South Carolina, Columbia, SC 29208, USA.
dickson@sc.edu

John Earman
Department of History and Philosophy of Science, 1017 Cathedral of Learning, University of Pittsburgh, Pittsburgh, PA 15260, USA.
jearman+@pitt.edu

George F. R. Ellis
Mathematics Department, University of Cape Town, Rondebosch, Cape Town 8001, South Africa.
ellis@maths.uct.ac.za

Gérard Emch
Department of Mathematics, University of Florida, 358 Little Hall, PO Box 118105, Gainesville, FL 32611–8105, USA.
gge@math.ufl.edu

Hans Halvorson
Department of Philosophy, Princeton University, Princeton, NJ 08544, USA.
hhalvors@princeton.edu

Gerard 't Hooft
Institute for Theoretical Physics, Utrecht University, Leuvenlaan 4, 3584 CC Utrecht, The Netherlands, and Spinoza Institute, Postbox 80.195, 3508 TD Utrecht, The Netherlands.
g.thooft@phys.uu.nl

N.P. Landsman
Institute for Mathematics, Astrophysics, and Particle Physics, Radboud Universiteit Nijmegen, Toernooiveld 1, 6525 ED Nijmegen, The Netherlands
landsman@math.ru.nl

David B. Malament
Department of Logic and Philosophy of Science, University of California at Irvine, 3151 Social Science Plaza, Irvine, CA 92697-5100, USA.
dmalamen@uci.edu

Michael Müger
Institute for Mathematics, Astrophysics, and Particle Physics, Radboud Universiteit Nijmegen, Toernooiveld 1, 6525 ED Nijmegen, The Netherlands.
mueger@math.ru.nl

Carlo Rovelli
Centre de Physique Théorique de Luminy, Université de la Méditerranée, case 907, F-13288 Marseille, France.
rovelli@cpt.univ-mrs.fr

Jos Uffink
Institute for History and Foundations of Science, Utrecht Univesrsity, PO Box 80 000, 3508 TA Utrecht, The Netherlands.
uffink@phys.uu.nl

COMPENDIUM OF THE FOUNDATIONS OF CLASSICAL STATISTICAL PHYSICS

Jos Uffink

1 INTRODUCTION

It has been said that an advantage of having a mature, formalized version of a theory is that one may forget its preceding history. This saying is certainly true for the purpose of studying the conceptual structure of a physical theory. In a discussion of the foundations of classical mechanics, for example, one need not consider the work of the Parisian scholastics. In the foundations of quantum mechanics, one may start from the von Neumann axioms, and disregard the preceding "old" quantum theory. Statistical physics, however, has not yet developed a set of generally accepted formal axioms, and consequently we have no choice but to dwell on its history.

This is not because attempts to chart the foundations of statistical physics have been absent, or scarce [Ehrenfest and Ehrenfest-Afanassjewa, 1912; ter Haar, 1955; Penrose, 1979; Sklar, 1993; Emch and Liu, 2001]e.g.. Rather, the picture that emerges from such studies is that statistical physics has developed into a number of different schools, each with its own programme and technical apparatus. Unlike quantum theory or relativity, this field lacks a common set of assumptions that is accepted by most of the participants; although there is, of course, overlap. But one common denominator seems to be that nearly all schools claim the founding fathers, Maxwell, Boltzmann and Gibbs as their champions.

Broadly understood, statistical physics may be characterized as a branch of physics intended to describe the thermal behaviour and properties of matter in bulk, i.e. of macroscopic dimensions in relation to its microscopic corpuscular constituents and their dynamics.[1] In this review, we shall only deal with approaches that assume a finite number of microscopic constituents, governed by classical dynamics. (See [Emch, 2006] for a discussion of quantum statistical physics that also addresses infinite systems.)

The above description is deliberately vague; it does not yet specify what thermal behaviour is, and being a characterization in terms of intentions, leaves open by what methods the goals may be achieved. Let us expand a bit. There are two basic

[1]The terms "in bulk" and the distinction "micro/macroscopic" should be understood in a relative sense. Thus, statistical physics may apply to a galaxy or nebula, in which the constituent stars are considered as 'microscopic constituents'.

Handbook of the Philosophy of Science. Philosophy of Physics
Volume editors: Jeremy Butterfield and John Earman
General editors: Dov M. Gabbay, Paul Thagard and John Woods
© 2007 Elsevier B.V. All rights reserved.

ingredients in statistical physics. The first is a mechanical model of a macroscopic material system. For example, a gas may be modeled as a system of point particles, or as hard spheres, or as composite objects, etc. Similarly, one may employ lattice models for solids, and so forth. In general, the particulars of the mechanical model, and its dynamics, will depend on the system of interest.

The second ingredient of the theory on which all approaches agree is the introduction of probability and statistical considerations. Sometimes, textbooks explain the need for this ingredient by pointing to the fact that an exact solution of the equations of motion for mechanical models with a large number of degrees of freedom is unfeasible. But this motivation from deficiency surely underestimates the constructive and explanatory role that probability plays in statistical physics. A slightly better motivation, also found in many textbooks, is that even if the dynamical equations could be solved in detail, most of these details would turn out to be irrelevant for the purpose of characterizing the thermal behaviour. There is some truth in this observation, yet it can hardly be satisfactory as it stands. Certainly, not all details about the microdynamics are irrelevant, e.g. in phase transitions, and one naturally wishes for more concrete information about exactly which details are irrelevant and which are not.

One of the foremost foundational problems in statistical physics is thus to specify and to clarify the status of probabilistic assumptions in the theory. As we shall see, this task already leads to a rough distinction between approaches in which probability arises as a notion explicitly defined in mechanical terms (kinetic theory), and approaches in which it is a conceptually independent ingredient (statistical mechanics).

Next, there are ingredients on which much less consensus can be found. Here is a (partial) list:

- Assumptions about the overwhelmingly large number of microscopic constituents (typically of the order of 10^{23} or more).

- An assumption about the erratic nature of the dynamics (e.g. ergodicity).

- The choice of special initial conditions.

- The role of external influences on the system, i.e., assumptions about whether the system is open to the exchange of energy/momentum with its environment, in combination with an assumed sensitivity of the dynamics under such external disturbances.

- Symmetry of macroscopic quantities under permutation of the microscopic constituents.

- Limits in the resolution or experimental accuracy of macroscopic observers.

- Appeal to a time-asymmetric principle of causality.

The role of each of these ingredients in the recipe of statistical physics is controversial. What many "chefs" regard as absolutely essential and indispensable, is argued to be insufficient or superfluous by many others. A major goal in the foundations of statistical physics should therefore lie in an attempt to sort out which subset of the above ideas can be formulated in a precise and coherent manner to obtain a unified and sufficiently general framework for a theory of statistical physics.

Another issue in which the preceding discussion has been vague is what is meant by the thermal behaviour and properties of macroscopic matter. There are two sources on which one may draw in order to delineate this topic. The first is by comparison to other (older) traditions in theoretical physics that have the same goal as statistical physics but do not rely on the two main ingredients above viz. a mechanical model and probabilistic arguments. There are two main examples: thermodynamics and hydrodynamics. The other source, of course, is observation. This provides a rich supply of phenomena, some of which have thus far withstood full theoretical explanation (e.g. turbulence).

Obviously, a measure of success for statistical physics can be found in the question to what extent this approach succeeds in reproducing the results of earlier, non-statistical theories, where they are empirically adequate, and in improving upon them where they are not. Thus, the foundations of statistical physics also provides a testing ground for philosophical ideas about inter-theory relations, like reduction (cf. [Brush, 1977; Sklar, 1993; Batterman, 2002]). However I will not go into this issue. The remainder of this introduction will be devoted to a rough sketch of the four theories mentioned, i.e. thermodynamics, hydrodynamics, kinetic theory and statistical physics.

1.1 Thermodynamics.

Orthodox thermodynamics is an approach associated with the names of Clausius, Kelvin, and Planck. Here, one aims to describe the thermal properties of macroscopic bodies while deliberately avoiding commitment to any hypothesis about the microscopic entities that might constitute the bodies in question. Instead, the approach aims to derive certain general laws, valid for all such bodies, from a restricted set of empirical principles.

In this approach the macroscopic body (or thermodynamic system) is conceived of as a sort of black box, which may interact with its environment by means of work and heat exchange. The most basic empirical principle is that macroscopic bodies when left to themselves, i.e. when isolated from an environment, eventually settle down in an equilibrium state in which no further observable changes occur. Moreover, for simple, homogeneous bodies, this equilibrium state is fully characterized by the values of a small number of macroscopic variables.

Other empirical principles state which types of processes are regarded as impossible. By ingenious arguments one can then derive from these principles the existence of certain quantities (in particular: absolute temperature, energy and entropy) as 'state functions', i.e. functions defined on a space of thermodynamical equilibrium states for all such systems.

While the theory focuses on processes, the description it can afford of such processes is extremely limited. In general, a process will take a system through a sequence of non-equilibrium states, for which the thermodynamic state functions are not defined, and thus cannot be characterized in detail with the tools afforded by the theory. Therefore one limits oneself to the consideration of special types of processes, namely those that begin and end in an equilibrium state. Even more special are those processes that proceed so delicately and slowly that up to an arbitrarily small error one may assume that the system remains in equilibrium throughout the entire process. The latter processes are called *quasistatic*, or sometimes *reversible*.[2]

Of course, since equilibrium states are by definition assumed to remain in equilibrium if unperturbed, all such processes are triggered by an external intervention such as pushing a piston or removing a partition. For the first type of process, orthodox thermodynamics can only relate the initial and final state. The second type of process can be (approximately) represented as a curve in the equilibrium state space.

The advantage of the approach is its generality. Though developed originally for the study of gases and liquids, by the late nineteenth century, it could be extended to the behaviour of magnets and other systems. Indeed, the independence of hypotheses about its micro-constituents means that the methods of orthodox thermodynamics can also be –and have been– applied to essentially quantum-mechanical systems (like photon gases) or to more exotic objects like black holes (see [Rovelli, 2006]).

With regard to the foundations of statistical physics, two aspects of thermodynamics are of outstanding importance. First, the challenge is to provide a counterpart for the very concept of equilibrium states and to provide a counterpart for the thermodynamic law that all isolated systems not in equilibrium evolve towards an equilibrium state. Secondly, statistical physics should give an account of the Second Law of thermodynamics, i.e. the statement that entropy cannot decrease in an adiabatically isolated system. Obviously, such counterparts will be statistical; i.e. they will hold on average or with high probability, but will not coincide with the unexceptionally general statements of thermodynamics.

1.2 Hydrodynamics

It would be a mistake to believe that the goals of statistical physics are exhausted by reproducing the laws of thermodynamics. There are many other traditions in theoretical physics that provide a much more detailed, yet less general, characterization of thermal behaviour. A concrete example is hydrodynamics or fluid dynamics. In contrast to thermodynamics, hydrodynamics does rely on an assump-

[2]The reader may be warned, however, that there are many different meanings to the term 'reversible' in thermodynamics. See [Uffink, 2001] for a discussion.

tion about microscopic constitution. It models a fluid as a continuous medium or plenum. It is, in modern parlance, a field theory. Moreover it aims to describe the evolution of certain macroscopic quantities in the course of time, i.e. during non-equilibrium processes. As such it is an example of a theory which is much more informative and detailed than thermodynamics, at the price, of course, that its empirical scope is restricted to fluids.

Without going in detail (for a more comprehensive account, see e.g. [Landau and Lifshitz, 1987; de Groot and Mazur, 1961]), hydrodynamics assumes there are three fundamental fields: the mass density $\rho(\vec{x}, t)$, a velocity field $\vec{v}(\vec{x}, t)$ and a temperature field $T(\vec{x}, t)$. There are also three fundamental field equations, which express, in a differential form, the conservation of mass, momentum and energy. Unfortunately, these equations introduce further quantities: the pressure $P(\vec{x}, t)$, the stress tensor $\pi(\vec{x}, t)$, the energy density $u(\vec{x}, t)$, the shear and bulk viscosities η and ζ and thermal conductivity κ, each of which has to be related to the fundamental fields by means of various constitutive relations and equations of state (dependent on the fluid concerned), in order to close the field equations, i.e. to make them susceptible to solution.

The resulting equations are explicitly asymmetric under time reversal. Yet another remarkable feature of hydrodynamics is the fact that the equations can be closed at all. That is: the specification of only a handful of macroscopic quantities is needed to predict the evolution of those quantities. Their behaviour is in other words *autonomous*. This same autonomy also holds for other theories or equations used to describe processes in systems out of equilibrium: for example the theories of diffusion, electrical conduction in metals, the Fourier heat equation etc. In spite of a huge number of microscopic degrees of freedom, the evolution of a few macroscopic quantities generally seems to depend only on the instantaneous values of these macroscopic quantities. Apart from accounting for the asymmetry under time reversal displayed by such theories, statistical physics should also ideally explain this remarkable autonomy of their evolution equations.

1.3 Kinetic theory

I turn to the second group of theories we need to consider: those that do rely on hypotheses or modeling assumptions about the internal microscopic constitution or dynamics of the systems considered. As mentioned, they can be divided into two rough subgroups: kinetic theory and statistical mechanics.

Kinetic theory, also called the kinetic theory of gases, the dynamical theory of gases, the molecular-kinetic theory of heat etc., takes as its main starting point the assumption that systems (gases in particular) consist of molecules. The thermal properties and behaviour are then related in particular to the motion of these molecules.

The earliest modern version of a kinetic theory is Daniel Bernoulli's (1731). Bernoulli's work was not followed by further developments along the same line for almost a century. But it regained new interest in the mid-nineteenth century.

The theory developed into a more general and elaborate framework in the hands of Clausius, Maxwell and Boltzmann. Clausius extended Bernoulli's model by taking into account the collisions between the particles, in order to show that the formidable molecular speeds (in the order of 10^3 m/s) were compatible with relatively slow rates of diffusion. However, he did not develop a systematic treatment of collisions and their effects. It was Maxwell who was the first to realize that collisions would tend to produce particles moving at a variety of speeds, rather than a single common speed, and proceeded to ask how probable the various values of the velocity would be in a state of equilibrium. Maxwell thus introduced the concept of probability and statistical considerations into kinetic theory.

From 1868 onwards, Boltzmann took Maxwell's investigations further. In his famous memoir of 1872 he obtained an equation for the evolution of the distribution function, the Boltzmann equation, and claimed that every non-stationary distribution function for an isolated gas would evolve towards the Maxwellian form, i.e. towards the equilibrium state. However, along the way, Boltzmann had made various assumptions and idealizations, e.g. neglecting the effect of multi-particle collisions, which restrict his derivations' validity to dilute gases, as well as the *Stoßzahlansatz*, developed by Maxwell in 1867, (or 'hypothesis of molecular disorder' as he later called it).

The Boltzmann equation, or variations of this equation, is the physicists' workhorse in gas theory. The hydrodynamical equations can be derived from it, as well as other transport equations. However, it is well known that it is only an approximation, and commonly regarded as a first step in a hierarchy of more detailed equations. But the foremost conceptual problem is its time-asymmetric nature, which highlights the fact that the Boltzmann equation itself could not be derived from mechanics alone. During Boltzmann's lifetime, this led to two famous objections, the reversibility objection (*Umkehreinwand*) by Loschmidt and the recurrence objection (*Wiederkehreinwand*) by Zermelo. A third important challenge, only put forward much more recently by [Lanford, 1975], concerns the consistency of the Boltzmann equation with the assumption that the gas system is a mechanical system governed by Hamiltonian dynamics.

1.4 Statistical mechanics

There is only a vague borderline between kinetic theory and statistical mechanics. The main distinctive criterion, as drawn by the Ehrenfests (1912) is this. Kinetic theory is what the Ehrenfests call "the older formulation of statistico-mechanical investigations" or "kineto-statistics of the molecule". Here, molecular states, in particular their velocities, are regarded as stochastic variables, and probabilities are attached to such molecular states of motion. These probabilities themselves are determined by the state of the total gas system. They are conceived of either as the relative *number* of molecules with a particular state, or the relative *time* during which a molecule has that state. (Maxwell employed the first option, Boltzmann wavered between the two.) It is important to stress that in both options the "probabilities" in question are determined by the *mechanical* properties of the gas. Hence there is really no clear separation between mechanical and statistical concepts in this approach.

Gradually, a transition was made to what the Ehrenfests called a "modern formulation of statistico-mechanical investigations" or "kineto-statistics of the gas model", or what is nowadays known as statistical mechanics. In this latter approach, probabilities are not attached to the state of a molecule but to the state of the entire gas system. Thus, the state of the gas, instead of determining the probability distribution, now itself becomes a stochastic variable.

A merit of this latter approach is that interactions between molecules can be taken into account. Indeed, the approach is not necessarily restricted to gases, but might in principle also be applied to liquids or solids. (This is why the name 'gas theory' is abandoned.) The price to be paid however, is that the probabilities themselves become more abstract. Since probabilities are attributed to the mechanical states of the total system, they are no longer determined by such mechanical states. Instead, in statistical mechanics, the probabilities are usually conceived of as being determined by means of an 'ensemble', i.e. a fictitious collection of replicas of the system in question. But whatever role one may wish to assign to this construction, the main point is that probability is now an independent concept, no longer reducible to mechanical properties of the system.

It is not easy to pinpoint this transition in the course of the history, except to say that Maxwell's work in the 1860s definitely belong to the first category, and Gibbs' book of 1902 to the second. Boltzmann's own works fall somewhere in the middle ground. His earlier contributions clearly belong to the kinetic theory of gases (although his 1868 paper already applies probability to an entire gas system); while his work after 1877 is usually seen as elements in the theory of statistical mechanics. However, Boltzmann himself never indicated a clear distinction between these two different theories, and any attempt to draw a demarcation at an exact location in his work seems somewhat arbitrary.

From a conceptual point of view, the transition from kinetic gas theory to statistical mechanics poses two main foundational questions. First: on what grounds do we choose a particular ensemble, or the probability distribution characterizing the ensemble? Gibbs did not enter into a systematic discussion of this problem, but only discussed special cases of equilibrium ensembles (i.e. canonical, microcanonical etc.) for which the probability distribution was stipulated by some special simple form. A second problem is to relate the ensemble-based probabilities to the probabilities obtained in the earlier kinetic approach for a single gas model.

The Ehrenfests [1912] paper was the first to recognize these questions, and to provide a partial answer. Namely: Assuming a certain hypothesis of Boltzmann's, which they dubbed the *ergodic hypothesis*, they pointed out that for an isolated system the micro-canonical distribution is the unique stationary probability distribution. Hence, if one demands that an ensemble of isolated systems describing thermal equilibrium must be represented by a stationary distribution, the only choice for this purpose is the micro-canonical one. Similarly, they pointed out that under the ergodic hypothesis, infinite time averages and ensemble averages were identical. This, then, would provide a desired link between the probabilities of the older kinetic gas theory and those of statistical mechanics, at least in equilibrium and in the infinite time limit. Yet the Ehrenfests simultaneously expressed strong doubts about the validity of the ergodic hypothesis. These doubts were soon substantiated when in 1913 Rosenthal and Plancherel proved that the hypothesis was untenable for realistic gas models.

The Ehrenfests' reconstruction of Boltzmann's work thus gave a prominent role to the ergodic hypothesis, suggesting that it played a fundamental and lasting role in his thinking. Although this view indeed produces a more coherent view of his multi-faceted work, it is certainly not historically correct. Boltzmann himself also had grave doubts about this hypothesis, and expressly avoided it whenever he could, in particular in his two great papers of 1872 and 1877b. Since the Ehrenfests, many authors have presented accounts of Boltzmann's work. Particularly important are Klein [1973] and Brush [1976].

Nevertheless, the analysis of the Ehrenfests did thus lead to a somewhat clearly delineated programme for or view about the foundations of statistical physics, in which ergodicity was a crucial feature. The demise of the original ergodic hypothesis did not halt the programme; the hypothesis was replaced by an alternative (weaker) hypothesis, i.e. that the system is '*metrically transitive*' (nowadays, the name 'ergodic' is often used as synonym). What is more, certain mathematical results of Birkhoff and von Neumann (the ergodic theorem) showed that for ergodic systems in this new sense, the desired results could indeed be proven, modulo a few mathematical provisos that at first did not attract much attention.

Thus there arose the ergodic or "standard" view on the foundations of statistical mechanics; (see, e.g. [Khinchin, 1949, p. 44]). On that view, the formalism of statistical mechanics emerges as follows: A concrete system, say a container with gas, is represented as a mechanical system with a very large number of degrees of freedom. All physical quantities are functions of the dynamical variables of the system, or, what amounts to the same thing, are functions on its phase space. However, experiments or observation of such physical quantities do not record the instantaneous values of these physical quantities. Instead, every observation must last a duration which may be extremely short by human standards, but will be extremely long on the microscopic level, i.e. one in which the microstate has experienced many changes, e.g. because of the incessant molecular collisions. Hence, all we can register are *time averages* of the physical quantities over a very long periods of time. These averages are thus empirically meaningful. Unfortunately they are theoretically and analytically obstreperous. Time averages depend on the trajectory and can only be computed by integration of the equations of motion. The expectation value of the phase function over a given ensemble, the *phase average* has the opposite qualities, i.e. it is easy to compute, but not immediately empirically relevant. However, ergodicity ensures that the two averages are equal (almost everywhere). Thus, one can combine the best of both worlds, and identify the theoretically convenient with the empirically meaningful.

While statistical mechanics is clearly a more powerful theory than kinetic theory, it is, like thermodynamics, particularly successful in explaining and modeling gases and other systems in equilibrium. Non-equilibrium statistical mechanics remains a field where extra problems appear.

1.5 Prospectus

The structure of this chapter is as follows. In Section 2, I will provide a brief exposition of orthodox thermodynamics, and in subsection 2.2 an even briefer review of some less-than-orthodox approaches to thermodynamics. Section 3 looks at the kinetic theory of gases, focusing in particular on Maxwell's ground-breaking papers of 1860 and 1867, and investigates the meaning and status of Maxwell's probabilistic arguments.

Section 4 is devoted to (a selection of) Boltzmann's works, which, as mentioned above, may be characterized as in between kinetic theory and statistical mechanics. The focus will be on his 1868 paper and his most celebrated papers of 1872 and 1877. Also, the objections from Loschmidt [1877] and Zermelo [1897] are discussed, together with Boltzmann's responses. Our discussion emphasizes the variety of assumptions and methods used by Boltzmann over the years, and the open-endedness of his results: the ergodic hypothesis, the *Stoßzahlansatz*, the combinatorial argument of 1877, and a statistical reading of the H-theorem that he advocated in the 1890s.

Next, Section 5 presents an account of Gibbs' [1902] version of statistical mechanics and emphasizes the essential differences between his and Boltzmann's approach. Sections 6 and 7 give an overview of some more recent developments in statistical mechanics, In particular, we review some results in modern ergodic theory, as well as approaches that aim to develop a more systematic account of non-equilibrium theory, such as the BBGKY approach (named after Bogolyubov, Born, Green, Kirkwood and Yvon) and the approach of Lanford. Section 7 extends this discussion for a combination of approaches, here united under the name *stochastic dynamics* that includes those known as 'coarse-graining' and 'interventionism' or 'open systems'. In all cases we shall look at the question whether or how such approaches succeed in a satisfactory treatment of non-equilibrium.

As this prospectus makes clear, the choice of topics is highly selective. There are many important topics and developments in the foundations of statistical physics that I will not touch. I list the most conspicuous of those here together with some references for readers that wish to learn more about them.

- Maxwell's demon and Landauer's principle: [Klein, 1970; Earman and Norton, 1998; 1999; Leff and Rex, 2003; Bennett, 2003; Norton, 2005; Maroney, 2005; Ladyman *et al.*, 2006].

- Boltzmann's work in the 1880s (e.g. on monocyclic systems) [Klein, 1972; 1974; Bierhalter, 1992; Gallavotti, 1999; Uffink, 2005].

- Gibbs' paradox [van Kampen, 1984; Jaynes, 1992; Huggett, 1999; Saunders, 2006].

- Branch systems [Schrödinger, 1950; Reichenbach, 1956; Kroes, 1985; Winsberg, 2004].

- Subjective interpretation of probability in statistical mechanics [Tolman, 1938; Jaynes, 1983; von Plato, 1991; van Lith, 2001a; Balian, 2005].

- Prigogine and the Brussels-Austin school [Obcemea and Brändas, 1983; Batterman, 1991; Karakostas, 1996; Edens, 2001; Bishop, 2004].

2 ORTHODOX THERMODYNAMICS

2.1 The Clausius-Kelvin-Planck approach

Thermodynamics is a theory that aims to characterize macroscopic physical bodies in terms of macroscopically observable quantities (typically: temperature, pressure, volume, etc.,) and to describe their changes under certain types of interactions (typically exchange of heat or work with an environment).

The classical version of the theory, which evolved around 1850, adopted as a methodological starting point that the fundamental laws of the theory should be independent of any particular hypothesis about the microscopic constitution of the bodies concerned. Rather, they should be based on empirical principles, i.e. boldly generalized statements of experimental facts, not on hypothetical and hence untestable assumptions such as the atomic hypothesis.

The reasons for this methodology were twofold. First, the dominant view on the goal of science was the positivist-empirical philosophy which greatly valued directly testable empirical statements above speculative hypotheses. But the sway of the positivist view was never so complete that physicists avoided speculation altogether. In fact many of the main founders of thermodynamics eagerly indulged in embracing particular hypotheses of their own about the microphysical constitution of matter.

The second reason is more pragmatic. The multitude of microphysical hypotheses and conjectures was already so great in the mid-nineteenth century, and the prospect of deciding between them so dim, that it was a clear advantage to obtain and present results that did not depend on such assumptions. Thus, when Clausius stated in 1857 that he firmly believed in the molecular-kinetic view on the nature of gases, he also mentioned that he had not previously revealed this opinion in order not to mix this conviction with his work on thermodynamics proper [Clausius, 1857, p. 353].[3]

[3]The wisdom of this choice becomes clear if we compare his fame to that of Rankine. Rankine actually predated Clausius in finding the entropy function (which he called 'thermodynamic potential'). However, this result was largely ignored due to the fact that it was imbedded in Rankine's rather complicated theory of atomic vortices.

Proceeding somewhat ahistorically,[4] one might say that the first central concept in thermodynamics is that of *equilibrium*. It is taken as a fact of experience that macroscopic bodies in a finite volume, when left to themselves, i.e. isolated from an environment eventually settle down in a stationary state in which no further observable changes occur (the 'Minus First Law', cf. page 939). This stationary state is called a (thermal) *equilibrium state*. Moreover, for simple, homogeneous bodies, this state is fully characterized by the values of a small number of macroscopic variables. In particular, for fluids (i.e. gases or liquids), two independent variables suffice to determine the equilibrium state.

For fluids, the three variables pressure p, temperature θ and volume V, are thus related by a so-called equation of state, where, following Euler, it has become customary to express pressure as a function of the two remaining variables:

$$(1) \quad p = p(\theta, V)$$

The form of this function differs for different fluids; for n moles of an ideal gas it is given by:

$$(2) \quad p(\theta, V) = nR\theta/V$$

where R is the gas constant and θ is measured on the gas thermometer scale.

The content of thermodynamics developed out of three ingredients. The first is the science of calorimetry, which was already developed to theoretical perfection in the eighteenth century, in particular by Joseph Black [Fox, 1971; Truesdell, 1980; Chang, 2003; 2004]. It involved the study of the thermal changes in a body under the addition of or withdrawal of heat to the system. Of course, the (silent) presupposition here is that this process of heat exchange proceeds so delicately and slowly that the system may always be regarded as remaining in equilibrium. In modern terms, it proceeds 'quasi-statically'. Thus, the equation of state remains valid during the process.

The tools of calorimetry are those of differential calculus. For an infinitesimal increment $đQ$ of heat added to a fluid, one puts

$$(3) \quad đQ = c_V d\theta + \Lambda_\theta dV,$$

where c_V is called the heat capacity at constant volume and Λ_θ the latent heat at constant temperature. Both c_V and Λ_θ are assumed to be functions of θ and V. The notation $đ$ is used to indicate that the heat increment $đQ$ is not necessarily an exact differential, i.e. Q is not assumed to be a function of state.

The total heat Q added to a fluid during a process can thus be expressed as a line integral along a path $\mathcal{P}$ in the (θ, V) plane

$$(4) \quad Q(\mathcal{P}) = \int_{\mathcal{P}} đQ = \int_{\mathcal{P}} (c_V d\theta + \Lambda_\theta dV)$$

[4] I refer to [Uffink, 2001] for more details.

A treatment similar to the above can be given for the quasistatic heat exchange of more general thermal bodies than fluids. Indeed, calorimetry was sufficiently general to describe phase transitions, say from water to ice, by assuming a discontinuity in Λ_θ.

All this is independent of the question whether heat itself is a substance or not. Indeed, Black himself wished to remain neutral on this issue. Even so, much of the terminology of calorimetry somehow invites the supposition that heat is a substance, usually called caloric, and many eighteenth and early nineteenth century authors adopted this view [Fox, 1971]. In such a view it makes sense to speak of the amount of heat contained in a body, and this would entail that $đQ$ must be an exact differential (or in other words: $Q(\mathcal{P})$ must be the same for all paths $\mathcal{P}$ with the same initial and final points). But this turned out to be empirically false, when the effects of the performance of work were taken into account.

Investigations in the 1840s (by Joule and Mayer among others) led to the conviction that heat and work are "equivalent"; or somewhat more precisely, that in every cyclic process $\mathcal{C}$, the amount of heat $Q(\mathcal{C})$ absorbed by the system is proportional to the amount of work performed by the system. Or, taking $W(\mathcal{C})$ as positive when performed *on* the system :

(5) $$JQ(\mathcal{C}) + W(\mathcal{C}) = 0$$

where $J \approx 4.2\text{Nm/Cal}$ is Joule's constant, which modern convention takes equal to 1. This is the so-called *First Law of thermodynamics.*

For quasistatic processes this can again be expressed as a line integral in a state space Ω_{eq} of thermodynamic equilibrium states

(6) $$\oint_{\mathcal{C}} (đQ + đW) = 0$$

where

(7) $$đW = -pdV.$$

Assuming the validity of (6) for all cyclic paths in the equilibrium state space implies the existence of a function U on Ω_{eq} such that

(8) $$dU = đQ + đW.$$

The third ingredient of thermodynamics evolved from the study of the relations between heat and work, in particular the efficiency of heat engines. In 1824, Carnot obtained the following theorem.

> CARNOT'S THEOREM: Consider any system that performs a cyclic process $\mathcal{C}$ during which (a) an amount of heat $Q_+(\mathcal{C})$ is absorbed from a heat reservoir at temperature θ_+, (b) an amount of heat $Q_-(\mathcal{C})$ is given off to a reservoir at a temperature θ_-, with $\theta_- < \theta_+$, (c) there is no heat exchange at other stages of the cycles, and (d) some work

$W(\mathcal{C})$ is done on a third body. Let $\eta(\mathcal{C}) := \frac{W(\mathcal{C})}{Q_+(\mathcal{C})}$ be the efficiency of the cycle. Then:

> (1) All quasistatic cycles have the same efficiency. This efficiency is a universal function of the two temperatures, i.e.,
>
> (9) $\eta(\mathcal{C}) = \eta(\theta_+, \theta_-)$.
>
> (2) All other cycles have a efficiency which is less or equal to that of the quasi-static cycle.

Carnot arrived at this result by assuming that heat was a conserved substance (and thus: $Q_+(\mathcal{C}) = Q_-(\mathcal{C})$ for all $\mathcal{C}$), as well as a principle that excluded the construction of a perpetuum mobile (of the first kind).

In actual fact, Carnot did not use the quasistatic/non-quasistatic dichotomy to characterize the two parts of his theorem. [5]

In fact, he used two different characterizations of the cycles that would produce maximum efficiency. (a): In his proof that Carnot cycles belong to class (1), the crucial assumption is that they "might have been performed in an inverse direction and order"[Carnot, 1824, p. 11]. But a little later (p. 13), he proposed a necessary and sufficient condition for a cycle to produce maximum efficiency, namely (b): In all stages which involve heat exchange, only bodies of equal temperature are put in thermal contact, or rather: their temperatures differ by a vanishingly small amount.

Carnot's theorem is remarkable since it did not need any assumption about the nature of the thermal system on which the cycle was carried out. Thus, when his work first became known to the physics community (Thomson, later known as Lord Kelvin, 1848) it was recognized as an important clue towards a general theory dealing with both heat and work exchange, for which Kelvin coined the name 'thermodynamics'. Indeed, Kelvin already showed in his first paper (1848) on the subject that Carnot's universal function η could be used to devise an absolute scale for temperature, i.e. one that did not depend on properties of a particular substance.

Unfortunately, around the very same period it became clear that Carnot's assumption of the conservation of heat violated the First Law. In a series of papers Clausius and Kelvin re-established Carnot's theorem on a different footing (i.e. on the first law (5) or, in this case $Q_+(\mathcal{C}) = Q_-(\mathcal{C}) + W(\mathcal{C})$, and a principle that excluded perpetual motion of the second kind) and transformed his results into general propositions that characterize general thermodynamical systems and their changes under the influence of heat and work. For the most part, these investigations were concerned with the first part of Carnot's theorem only. They led to what is nowadays called the first part of the Second Law; as follows.

First, Kelvin reformulated his 1848 absolute temperature scale into a new one, $T(\theta)$, in which the universal efficiency could be expressed explicitly as:

---

[5] Indeed, [Truesdell, 1980] argues that this characterization of his theorem is incorrect. See [Uffink, 2001] for further discussions.

(10) $\eta(T_+, T_-) = 1 - \frac{T_-}{T_+}$,

where $T_i = T(\theta_i)$. Since the efficiency η is also expressed by $W/Q_+ = 1-(Q_-/Q_+)$, this is equivalent to

(11) $\frac{Q_-}{T_-} = \frac{Q_+}{T_+}$.

Next, changing the sign convention to one in which Q is positive if absorbed and negative if given off by the system, and generalizing for cycles in which an arbitrary number of heat reservoirs are involved, one gets:

(12) $\sum_i \frac{Q_i}{T_i} = 0$.

In the case where the system is taken through a quasistatic cycle in which the heat reservoirs have a continuously varying temperature during this cycle, this generalizes to

(13) $\oint_{\mathcal{C}} \frac{\text{đ}Q}{T} = 0$.

Here, T still refers to the temperature of the heat reservoirs with which the system interacts, not to its own temperature. Yet Carnot's necessary and sufficient criterion of reversibility itself requires that during all stages of the process that involve heat exchange, the temperatures of the heat reservoir and system should be equal. Hence, in this case one may equate T with the temperature of the system itself.

The virtue of this result is that the integral (13) can now be entirely expressed in terms of quantities of the system. By a well-known theorem, applied by Clausius in 1865, it follows that there exists a function, called entropy S, defined on the equilibrium states of the system such that

(14) $S(s_1) - S(s_2) = \int_{s_1}^{s_2} \frac{\text{đ}Q}{T}$

or, as it more usually known:

(15) $\frac{\text{đ}Q}{T} = dS$.

This result is frequently expressed as follows: $\text{đ}Q$ has an *integrating divisor* (namely T): division by T turns the inexact (incomplete, non-integrable) differential $\text{đ}Q$ into an exact (complete, integrable) differential. For one mole of ideal gas (i.e. a fluid for which c_V is constant, Λ_θ vanishes and the ideal gas law (2) applies), one finds, for example:

(16) $S(T, V) = c_V \ln T + R \ln V + const.$

The existence of this entropy function also allows for a convenient reformulation of the First Law for quasistatic processes (8) as

(17) $dU = TdS - pdV,$

now too expressed in terms of properties of the system of interest.

However important this first part of the Second Law is by itself, it never led to much dispute or controversy. By contrast, the extension of the above results to cover the second part of Carnot's theorem gave rise to considerably more thought, and depends also intimately on what is understood by '(ir)reversible processes'.

The second part of Carnot's theorem was at first treated in a much more stepmotherly fashion. Clausius' [1854] only devoted a single paragraph to it, obtaining the result that for "irreversible" cycles

(18) $\oint \frac{dQ}{T} \leq 0.$

But this result is much less easy to apply, since the temperature T here refers to that of the heat reservoir with which the system is in contact, not (necessarily) that of the system itself.

Clausius put the irreversible processes in a more prominent role in his 1865 paper. If an irreversible cyclic process consists of a general, i.e. possibly non-quasistatic stage, from s_i to s_f, and a quasistatic stage, from s_f back to s_i, one may write (18) as

(19) $\int_{s_i\ \mathrm{non-qs}}^{s_f} \frac{đQ}{T} + \int_{s_f\ \mathrm{qs}}^{s_i} \frac{đQ}{T} \leq 0.$

Applying (14) to the second term in the left hand side, one obtains

(20) $\int_{s_i\ \mathrm{non-qs}}^{s_f} \frac{đQ}{T} \leq S(s_f) - S(s_i)$

If we assume moreover that the generally non-quasistatic process is adiabatic, i.e. $đQ = 0$, the result is

(21) $S(s_i) \leq S(s_f).$

In other words, in any adiabatic process the entropy of the final state cannot be less than that of the initial state.

Remarks: 1. The notation $\oint$ for cyclic integrals, and $đ$ for inexact differentials is modern. Clausius, and Boltzmann after him, would simply write $\int \frac{dQ}{T}$ for the left-hand side of (13) and (18).

2. An important point to note is that Clausius' formulation of the Second Law, strictly speaking, does not require a general monotonic increase of entropy for any adiabatically isolated system in the course of time. Indeed, in orthodox

thermodynamics, entropy is defined only for equilibrium states. Therefore it is meaningless within this theory to ask how the entropy of a system changes during a non-quasistatic process. All one can say in general is that when a system starts out in an equilibrium state, and ends, after an adiabatic process, again in an equilibrium state, the entropy of the latter state is not less than that of the former.

Still, the Second Law has often been understood as demanding continuous monotonic increase of entropy in the course of time, and often expressed, for adiabatically isolated systems, in a more stringent form

$$(22)\quad \frac{dS}{dt} \geq 0.$$

There is, however, no basis for this demand in orthodox thermodynamics.

3. Another common misunderstanding of the Second Law is that it would only require the non-decrease of entropy for processes in *isolated* systems. It should be noted that this is only part of the result Clausius derived: the Second Law holds more generally for *adiabatic* processes, i.e., processes during which the system remains adiabatically insulated. In other words, the system may be subject to arbitrary interactions with the environment, except those that involve heat exchange. (For example: stirring a liquid in a thermos flask, as in Joule's 'paddle wheel' experiment.)

4. Another point to be noted is that Clausius' result that the entropy in an adiabatically isolated system can never decrease is derived from the *assumption* that one can find a quasistatic process that connects the final to the initial state, in order to complete a cycle. Indeed, if such a process did not exist, the entropy difference of these two states would not be defined. The existence of such quasistatic processes is not problematic in many intended applications (e.g. if s_f and s_i are equilibrium states of a fluid); but it may be far from obvious in more general settings (for instance if one considers processes far from equilibrium in a complex system, such as a living cell). This warning that the increase of entropy is thus conditional on the existence of quasistatic transitions has been pointed out already by [Kirchhoff, 1894, p. 69].

5. Apart from the well-known First and Second Laws of thermodynamics, later authors have identified some more basic assumptions or empirical principles in the theory that are often assumed silently in traditional presentations — or sometimes explicitly but unnamed — which may claim a similar fundamental status.

The most familiar of these is the so-called *Zeroth Law*, a term coined by [Fowler and Guggenheim, 1939]. To introduce this, consider the *relation of thermal equilibrium.* This is the relationship holding between the equilibrium states of two systems, whenever it is the case that the composite system, consisting of these two systems, would be found in an equilibrium state if the two systems are placed in direct thermal contact — i.e., an interaction by which they are only allowed to exchange heat. The zeroth law is now that the assumption that this is a *transitive* relationship, i.e. if it holds for the states of two bodies A and B, and also for the states of bodies B and C, it likewise holds for bodies A and C.[6]

2.2 Less orthodox versions of thermodynamics

Even within the framework of orthodox thermodynamics, there are approaches that differ from the Clausius-Kelvin-Planck approach. The foremost of those is undoubtedly the approach developed by Gibbs in 1873–1878 [Gibbs, 1906]. Gibbs' approach differs much in spirit from his European colleagues. No effort is devoted to relate the existence or uniqueness of the thermodynamic state variables U T or S to empirical principles. There existence is simply assumed. Also, Gibbs focused on the description of equilibrium states, rather than processes.

Previous authors usually regarded the choice of variables in order to represent a thermodynamic quantity as a matter of convention, like the choice of a coordinate system on the thermodynamic (equilibrium) state space. For a fluid, one could equally well choose the variables (p, V), (V, T), etc., as long as they are independent and characterize a unique thermodynamic equilibrium state.[7] Hence one could equally well express the quantities U, S, etc. in terms of any such set of variables. However, Gibbs had the deep insight that some choices are 'better' than others, in the sense that if, e.g., the entropy is presented as a function of energy and volume, $S(U, V)$, (or energy as a function of entropy and volume, $U(S, V)$) all other thermodynamic quantities could be determined from it, while this is generally not true for other choices. For example, if one knows only that for one mole of gas $S(T, V)$ is given by (2), one cannot deduce the equations of state $p = RT/V$ and $U = c_V T$. In contrast, if the function $S(U, V) = c_V \ln U + R \ln V + \text{const.}$' is given, one obtains these equations from its partial derivatives: $\frac{p}{T} = (\frac{\partial S}{\partial V})_U$ and $\frac{1}{T} = (\frac{\partial S}{\partial U})_V$.

For this reason, Gibbs called

$$(23) \quad dU = TdS - pdV \quad \text{or} \quad dS = \frac{1}{T}dU + \frac{p}{T}dV$$

the *fundamental equation*.[8] Of course this does not mean that other choices of variables are inferior. Instead, one can find equivalent fundamental equations for

[6] Actually, transitivity alone is not enough. The assumption actually needed is that thermal equilibrium is an *equivalence* relation, i.e., it is transitive, reflective and symmetric (cf. [Boyling, 1972, p. 45]). The idea of elevating this to a fundamental 'Law', is that this assumption, which underlies the concept of temperature, can only be motivated on empirical grounds.

Another such assumption, again often stated but rarely named, is that any system contained in a finite volume, if left to itself, tends to evolve towards an equilibrium state. This has also sometimes been called a 'zeroth law' (cf. [Uhlenbeck and Ford, 1963, p.5; Lebowitz, 1994, p. 135]) in unfortunate competition with Fowler & Guggenheim's nomenclature. The name *Minus First Law* has therefore been proposed by [Brown and Uffink, 2001]. Note that this assumption already introduces an explicitly time-asymmetric element, which is deeper than — and does not follow from — the Second Law. However, most nineteenth (and many twentieth) century authors did not appreciate this distinction, and as we shall see below, this Minus First Law is often subsumed under the Second Law.

[7] The latter condition may well fail: A fluid like water can exist at different equilibrium states with the same p, V, but different T [Thomsen and Hartka, 1962]

[8] Note how Gibbs' outlook differs here from the Clausius-Kelvin-Planck view: These authors

such pairs of variables too, in terms of the Legendre transforms of U. (Namely: the Helmholtz free energy $F = U - TS$ for the pair (T, V); the enthalpy $U + pV$ for (p, S), and the Gibbs free energy $U + pV - TS$ for (p, T).) Further, Gibbs extended these considerations from homogeneous fluids to heterogeneous bodies, consisting of several chemical components and physical phases.

Another major novelty is that Gibbs proposed a variational principle to distinguish stable from neutral and unstable equilibria. (Roughly, this principle entails that for stable equilibrium the function $S(U, V)$ should be concave.) This criterium serves to be of great value in characterizing phase transitions in thermodynamic systems, e.g. the Van der Waals gas (Maxwell used it to obtain his famous "Maxwell construction" or equal area rule [Klein, 1978]). Gibbs work also proved important in the development of chemical thermodynamics, and physical chemistry.

Another group of approaches in orthodox thermodynamics is concerned particularly with creating a more rigorous formal framework for the theory. This is often called *axiomatic thermodynamics*. Of course, choosing to pursue a physical theory in an axiomatic framework does not by itself imply any preference for a choice in its physical assumptions or philosophical outlook. Yet the Clausius-Kelvin-Planck approach relies on empirical principles and intuitive concepts that may seem clear enough in their relation to experience — but are often surprisingly hard to define. Hence, axiomatic approaches tend to replace these empirical principles by statements that are conceptually more precise, but also more abstract, and thus arguably further removed from experience. The first example of this work is Carathéodory 1909. Later axiomatic approaches were pursued, among others, by [Giles, 1964]; [Boyling, 1972]; Jauch [1972; 1975], and by [Lieb and Yngvason, 1999]. All these approaches differ in their choice of primitive concepts, in the formulation of their axioms, and hence also in the results obtained and goals achieved. However, in a rough sense, one might say they all focus particularly on demonstrating under what conditions one might guarantee the mathematical existence and uniqueness of entropy and other state functions within an appropriate structure.

Since the 1940s a great deal of work has been done on what is known as "*non-equilibrium thermodynamics*" or "thermodynamics of irreversible processes" (see e.g. [de Groot, 1951; Prigogine, 1955; de Groot and Mazur, 1961, Yourgrau, 1966; Truesdell, 1969; Müller, 2003]). This type of work aims to extend orthodox thermodynamics into the direction of a description of systems in non-equilibrium states. Typically, one postulates that thermodynamic quantities are represented as continuously variable fields in space and time, with equilibrium conditions holding approximately within each infinitesimal region within the thermodynamic system.

would look upon (23) as a statement of the first law of thermodynamics, interpreting the differentials as infinitesimal increments during a quasistatic *process*, cf. (17). For Gibbs, on the other hand, (23) does not represent a process but a differential equation on the thermodynamic state space whose solution $U(S, V)$ or $S(V, U)$ contains *all* information about the equilibrium properties of the system, including the equations of state, the specific and latent heat, the compressibility, etc. — much more than just First Law.

Again, it may be noted that workers in the field seem to be divided into different schools (using names such as "extended thermodynamics","generalized thermodynamics", "rational thermodynamics", etc.) that do not at all agree with each other (see [Hutter and Wang, 2003]).

This type of work has produced many successful applications. But it seems fair to say that until now almost all attention has gone to towards practical application. For example, questions of the type that axiomatic thermodynamics attempts to answer, (e.g.: Under what conditions can we show the existence and uniqueness of the non-equilibrium quantities used in the formalism?) are largely unanswered, and indeed have given rise to some scepticism (cf. [Meixner, 1969; Meixner, 1970]). Another inherent restriction of this theory is that by relying on the assumption that non-equilibrium states can, at least in an infinitesimal local region, be well approximated by an equilibrium state, the approach is incapable of encompassing systems that are very far from equilibrium, such as in turbulence or living cells.)

The final type of approach that ought to be mentioned is that of *statistical thermodynamics*. The basic idea here is that while still refraining from introducing hypotheses about the microscopic constituents of thermodynamic systems, one rejects a key assumption of orthodox thermodynamics, namely, that a state of equilibrium is one in which all quantities attain constant values, in order to accommodate fluctuation phenomena such as Brownian motion or thermal noise. Thus the idea becomes to represent at least some of the thermodynamic quantities as random quantities, that in the course of time attain various values with various probabilities. Work in this direction has been done by Szilard [1925], Mandelbrot [1956; 1962; 1964], and Tisza and Quay [1963].

Of course the crucial question is then how to choose the appropriate probability distributions. One approach, elaborated in particular by [Tisza, 1966], taking its inspiration from Einstein [1910], relies on a inversion of Boltzmann's principle: whereas Boltzmann argued (within statistical mechanics) that the thermodynamic notion of entropy could be identified with the logarithm of a probability; Einstein argued that in thermodynamics, where the concept of entropy is already given, one may define the relative probability of two equilibrium states by the exponent of their entropy difference. Other approaches have borrowed more sophisticated results from mathematical statistics. For example, Mandelbrot used the Pitman-Koopman-Darmois theorem, which states that sufficient estimators exist only for the "exponential family" of probability distributions to derive the canonical probability distribution from the postulate that energy be a sufficient estimator of the system's temperature (see also [Uffink and van Lith, 1999]).

3 KINETIC THEORY FROM BERNOULLI TO MAXWELL

3.1 Probability in the mid-nineteenth century

Probability theory has a history dating back at least two centuries before the appearance of statistical physics. Usually, one places the birth of this theory in the correspondence of Pascal and Fermat around 1650. It was refined into a mature mathematical discipline in the work of Jacob Bernoulli [1713], Abraham de Moivre [1738] and Pierre-Simon de Laplace [1813] (cf. [Hacking, 1975]).

In this tradition, often called 'classical probability', the notion of probability is conceived of as a measure of the degree of certainty of our beliefs. Two points are important to note here. First, in this particular view, probability resides in the mind. There is nothing like uncertainty or chance in Nature. In fact, all authors in the classical tradition emphasize their adherence to strict determinism, either by appeal to divine omniscience (Bernoulli, de Moivre) or by appeal to the laws of mechanics and the initial conditions (Laplace). A probability hence represents a judgment about some state of affairs, and not an intrinsic property of this state of affairs. Hence, the classical authors never conceived that probability has any role to play in a description of nature or physical processes as such.[9] Secondly, although Bernoulli himself used the term "subjective" to emphasize the fact that probability refers to us, and the knowledge we possess, the classical interpretation does not go so far as modern adherents to a subjective interpretation of probability who conceive of probability as the degrees of belief of an arbitrary (although coherent) person, who may base his beliefs on personal whims, prejudice and private opinion.

This classical conception of probability would, of course, remain a view without any bite, if it were not accompanied by some rule for assigning values to probabilities in specific cases. The only such available rule is the so-called 'principle of insufficient reason': whenever we have no reason to believe that one case rather than another is realized, we should assign them equal probabilities (cf. [Uffink, 1995]). A closely related version is the rule that two or more variables should be independent whenever we have no reason to believe that they influence each other.

While the classical view was the dominant, indeed the only existent, view on probability for the whole period from 1650 to 1813, it began to erode around 1830. There are several reasons for this, but perhaps the most important is, paradoxically, the huge success with which the theory was being applied to the most varied subjects. In the wake of Laplace's influential *Essai philosophique sûr les Probabilités*, scientists found applications of probability theory in jurisdiction, demography, social science, hereditary research, etc. In fact, one may say: almost everywhere except physics (cf. [Hacking, 1990]). The striking regularity found in the frequencies of mass phenomena, and observations that (say) the number of raindrops per second on a tile follows the same pattern as the number of soldiers in the Prussian army killed each year by a kick from their horse, led to the alternative view that probability was not so much a representation of subjective (un)certainty, but rather the expression of a particular regularity in nature (Poisson, Quetelet). From these days onward we find mention of the idea of *laws of probability*, i.e. the idea that theorems of probability theory reflect lawlike behaviour to which

[9]Daniel Bernoulli might serve as an example. He was well acquainted with the work on probability of his uncle Jacob and, indeed, himself one of the foremost probabilists of the eighteenth century. Yet, in his work on kinetic gas theory (to be discussed in section 3.2), he did not find any occasion to draw a connection between these two fields of his own expertise.

Nature adheres. In this alternative, frequentist view of probability, there is no obvious place for the principle of insufficient reason. Instead, the obvious way to determine the values of probabilities is to collect empirical data on the frequencies on occurrences of events. However, a well-articulated alternative to the classical concept of probability did not emerge before the end of the century, and (arguably) not before 1919 — and then within in a few years there were no less than three alternatives: a logical interpretation by Keynes, a frequentist interpretation by von Mises and a subjective interpretation by Ramsey and De Finetti. See [Fine, 1973], [Galavotti, 2004] or [Emch, 2005] for a more detailed exposition.

Summing up roughly, one may say that around 1850 the field of probability was in a state of flux and confusion. Two competing viewpoints, the classical and the frequency interpretation, were available, and often mixed together in a confusing hodgepodge. The result was well-characterized in a famous remark of [Poincaré, 1896] that all mathematicians seem to believe that the laws of probability refer to statements learned from experience, while all natural scientists seem to think they are theorems derived by pure mathematics.

The work of Maxwell and Boltzmann in the 1860s emerged just in the middle of this confusing era. It is only natural that their work should reflect the ambiguity that the probability concept had acquired in the first half of the nineteenth century. Nevertheless, it seems that they mainly thought of probability in terms of frequencies, as an objective quantity, which characterizes a many-particle system, and that could be explicitly defined in terms of its mechanical state. This, however, is less clear for Maxwell than for Boltzmann.

Gradually, probability was emancipated from this mechanical background. Some isolated papers of Boltzmann [1871b] and Maxwell [1879] already pursued the idea that probabilities characterize an *ensemble* of many many-particle systems rather than a single system. Gibbs's 1902 book adopted this as a uniform coherent viewpoint. However, this ensemble interpretation is still sufficiently vague to be susceptible to different readings. A subjective view of ensembles, closely related to the classical interpretation of Bernoulli and Laplace, has emerged in the 1950s in the work of Jaynes. This paper, will omit a further discussion of this approach. I refer to [Jaynes, 1983; Uffink, 1995; 1996; Balian, 2005] for more details.

3.2 From Bernoulli to Maxwell (1860)

The kinetic theory of gases (sometimes called: dynamical theory of gases) is commonly traced back to a passage in Daniel Bernoulli's *Hydrodynamica* of 1738. Previous authors were, of course, quite familiar with the view that gases are composed of a large but finite number of microscopic particles. Yet they usually explained the phenomenon of gas pressure by a static model, assuming repulsive forces between these particles.

Bernoulli's discussion is the first to explain pressure as being due to their motion. He considered a gas as consisting of a great number of particles, moving hither and thither through empty space, and exerting pressure by their incessant collisions on the walls. With this model, Bernoulli was able to obtain the ideal gas law $pV = const.$ at constant temperature, predicted corrections to this law at high densities, and argued that the temperature could be taken as proportional to the square of the velocity of the particles. Despite this initial success, no further results were obtained in kinetic gas theory during the next century. By contrast, the view that modeled a gas as a continuum proved much more fertile, since it allowed the use of powerful tools of calculus. Indeed, the few works in the kinetic theory in the early nineteenth century e.g. by Waterston and Herapath were almost entirely ignored by their contemporaries (cf. [Brush, 1976]).

Nevertheless, the kinetic view was revived in the 1850s, in works by Kronig and Clausius. The main stimulus for this revival was the Joule-Mayer principle of the equivalence of heat and work, which led to the First Law of thermodynamics, and made it seem more plausible that heat itself was just a form of motion of gas particles. (A point well-captured in the title of Clausius' 1857 paper: "The kind of motion we call heat", subsequently adopted by Stephen Brush 1976 for his work on the history of this period.)

Clausius also recognized the importance of mutual collisions between the particles of the gas, in order to explain the relative slowness of diffusion when compared with the enormous speed of the particles (estimated at values of 400 m/s or more at ordinary room temperature). Indeed, he argued that in spite of their great speed, the mean free path, i.e. the distance a particle typically travels between two collision, could be quite small (of the order of micrometers) so that the mean displacement per second of particles is accordingly much smaller.

However, Clausius did not pay much attention to the consideration that such collisions would also change the magnitude of the velocities. Indeed, although his work sometimes mentions phrases like "mean speed" or "laws of probability" he does not specify a precise averaging procedure or probability assumption, and his calculations often proceed by crude simplifications (e.g. assuming that all but one of the particles are at rest).

Maxwell (1860)

It was Maxwell's paper of 1860 that really marks the re-birth of kinetic theory. Maxwell realized that if a gas consists of a great number N of moving particles, their velocities will suffer incessant change due to mutual collisions, and that a gas in a stationary state should therefore consist of a mixture of slower and faster particles. More importantly, for Maxwell this was not just an annoying complication to be replaced by simplifying assumptions, but the very feature that deserved further study.

He thus posed the question

> Prop. IV. To find the average number of particles whose velocities lie between given limits, after a great number of collisions among a great number of equal particles. [Maxwell, 1860, p. 380].

Denoting this desired average number as $Nf(\vec{v})d^3\vec{v}$, he found a solution to this problem by imposing two assumptions: the distribution function $f(\vec{v})$ should (i) factorize into functions of the orthogonal components of velocity, i.e. there exists some function g such that:

(24) $f(\vec{v}) = g(v_x)g(v_y)g(v_z)$,

and (ii) be spherically symmetric, i.e.,

(25) $f(\vec{v})$ depends only on $v = \|\vec{v}\|$.

He observed that these functional equations can only be satisfied if

(26) $f(\vec{v}) = Ae^{-v^2/B}$,

where the constant A is determined by normalization: $A = (B\pi)^{-3/2}$; and constant B is determined by relating the mean squared velocity to the absolute temperature — i.e., adopting modern notation: $\frac{3}{2}kT = \frac{m}{2}\langle v^2 \rangle$ — to obtain:

(27) $f(\vec{v}) = \left(\dfrac{m}{2\pi kT}\right)^{3/2} e^{-mv^2/2kT}$.

Maxwell's result led to some novel and unexpected predictions, the most striking being that the viscosity of a gas should be independent of its density, which was, nevertheless, subsequently experimentally verified. Another famous prediction of Maxwell was that in this model the ratio of the specific heats $\gamma = \frac{c_V}{c_p}$ must take the value of $\frac{4}{3}$. This did not agree with the experimentally obtained value of $\gamma = 1.408$.[10]

Maxwell's paper is the first to characterize the state of a gas by a distribution function f. It is also the first to call $f(\vec{v})d^3\vec{v}$ a *probability*. Clearly, Maxwell adopted a frequency interpretation of probability. The probability for the velocity to lie within a certain range $d^3\vec{v}$ is nothing but the relative number of particles in the gas with a velocity in that range. It refers to an objective, mechanical property of the gas system, and not to our opinions.

Now an obvious problem with this view is that if the gas contains a finite number of particles, the distribution of velocities must necessarily be discrete, i.e., in Dirac delta notation:

(28) $f(\vec{v}) = \dfrac{1}{N}\displaystyle\sum_{i=1}^{N} \delta(\vec{v} - \vec{v}_i)$,

[10]More generally, $c_V/c_p = (f+2)/f$ where f is the number of degrees of freedom of a molecule. This so-called c_V/c_p anomaly haunted gas theory for another half century. The experimental value around 1.4 is partly due to the circumstance that most ordinary gases have diatomic molecules for which, classically, $f = 6$. Quantum theory is needed to explain that one of these degrees is "frozen" at room temperature. Experimental agreement with Maxwell's prediction was first obtained by Kundt and Warburg in 1875 for mercury vapour. (For more details, see [Brush, 1976, p. 353–356]).

and if the energy of the gas is finite and fixed, the distribution should have a bounded support. The function (26) has neither of these properties.

It is not clear how Maxwell would have responded to such problems. It seems plausible that he would have seen the function (26) as representing only a good enough approximation,[11] in some sense, to the actual state of the gas but not to be taken too literally, just like actual frequencies in a chance experiment never match completely with their expected values. This is suggested by Maxwell's own illustration of the continuous distribution function as a discrete cloud of points, each of which representing the endpoint of a velocity vector (cf. Fig. 1 from [Maxwell, 1875]). This suggests he thought of an actual distribution more along the lines of (28) than (26). But this leaves the question open in what sense the Maxwell distribution approximates the actual distribution of velocities.

One option here would be to put more emphasis on the phrase "average" in the above quote from Maxwell. That is, maybe f is not intended to represent an actual distribution of velocities but an averaged one. But then, what kind of average? Since an average over the particles has already been performed, the only reasonable options could be an average over time or averaging over an ensemble of similar gas systems. But I can find no evidence that Maxwell conceived of such procedures in this paper. Perhaps one might argue that the distribution (26) is intended as an expectation, i.e. that it represents a reasonable mind's guess about the number of particles with a certain velocity. But in that case, Maxwell's interpretation of probability ultimately becomes classical.

However this may be, it is remarkable that the kinetic theory was thus able to make progress beyond Bernoulli's work by importing mathematical methods (functional equations) involving the representation of a state by continuous functions; though at the price of making this state concept more abstractly connected to physical reality.

A more pressing problem is that the assumptions (24, 25) Maxwell used to derive the form of his distribution do not sit well with its intended frequency interpretation. They seem to reflect a priori desiderata of symmetry, and are perhaps motivated by an appeal to some form of the principle of insufficient reason, in the sense that if there is, in our knowledge, no reason to expect a dependence between the various orthogonal components of velocity, we are entitled to assume they are independent.

This reading of Maxwell's motivations is suggested by the fact that in 1867 he described his 1860 assumption (24) as "the assumption that the probability of a molecule having a velocity resolved parallel to x lying between given limits is not in any way affected by the *knowledge* that the molecule has a given velocity resolved parallel to y" [Maxwell, 1867, emphasis added].

It has been pointed out (see e.g. [Brush, 1976, Vol. II, pp. 183–188]) that Maxwell's 1860 argument seems to have been heavily inspired by [Herschel, 1850] review

[11]This view was also expressed by [Boltzmann, 1896b]. He wrote, for example: "For a finite number of molecules the Maxwell distribution can never be realized exactly, but only as a good approximation" [Boltzmann, 1909, III, p. 569].

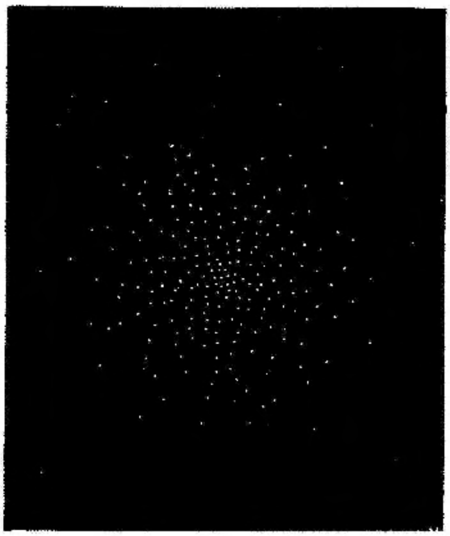


Diagram of Velocities.

Figure 1. An illustration of the Maxwell distribution from [Maxwell, 1875]. Every dot represents the end-point of a velocity vector.

of Quetelet's work on probability. This review essay contained a strikingly similar argument, applied to a marksman shooting at a target, in order to determine the probability that a bullet will land at some distance from the target. What is more, Herschel's essay is firmly committed to the classical interpretation of probability and gives the principle of insufficient reason a central role. Indeed, he explains the (analogue of) condition (25) as "nothing more than the expression of our state of *complete* ignorance of the causes of the errors [i.e. the deviation from the target] and their mode of action" [Herschel, 1850, p. 398, emphasis in the original]. If Maxwell indeed borrowed so much from Herschel, it seems likely that he would also have approved of, or at least be inspired by, this motivation of condition (25).[12]

Whatever may have been Maxwell's original motivation for these assumptions, their dubious nature is also clear from the fact that, in spite of his formulation of the problem (i.e. to determine the form of the function f "after a great number of collisions"), they do not refer to collisions at all. Indeed, it would seem that any motivation for their validity would just as well apply to a gas model consisting of non-colliding (e.g. perfectly transparent) particles as well. As Maxwell himself later remarked about certain derivations in the works of others, one might say that the condition "after a great number of collisions" is intended "rather for the sake of enabling the reader to form a mental image of the material system than as a condition for the demonstration" ([Garber *et al.*, 1995, p. 359]. [Maxwell, 1879]).

3.3 Maxwell (1867)

Whatever the merits and problems of his first paper, Maxwell's next paper on gas theory of 1867 rejected his previous attempt to derive the distribution function from the assumptions (24, 25) as "precarious" and proposed a completely different argument. This time, he considered a model of point particles with equal masses interacting by means of a repulsive central force, proportional to the fifth power of their mutual distance. What is more important, this time the collisions are used in the argument.

Maxwell considers an elastic collision between a pair of particles such that the initial velocities are $\vec{v}_1, \vec{v}_2$ and final velocities $\vec{v}_1{}', \vec{v}_2{}'$).[13] These quantities are related by the conservation laws of momentum and energy, yielding four equations, and two parameters depending on the geometrical factors of the collision process.

It is convenient to consider a coordinate frame such that particle 1 is at rest in the origin, and the relative velocity $\vec{v}_2 - \vec{v}_1$ is directed along the negative z axis, and

[12] It is interesting to note that Herschel's review prompted an early and biting criticism of the principle of insufficient reason as applied to frequencies of events by Leslie Ellis, containing the famous observation: "Mere ignorance is no ground for any inference whatsoever. *Ex nihilo nihil.* It cannot be that because we are ignorant of the matter, we know something about it" [Ellis, 1850]. It is not certain, however, whether Maxwell knew of this critique.

[13] In view of the infinite range of the interaction, 'initial' and 'final' are to be understood in an asymptotic sense, i.e. in the limits $t \longrightarrow \pm\infty$. An alternative followed in the text is to replace Maxwell's (1867) model with the hard spheres he had considered in 1860.

to use cylindrical coordinates. If (b, ϕ, z) denote the coordinates of the trajectory of the centre of particle 2, we then have $b = const.$, $\phi = const$, $z(t) = z_0 - \|\vec{v}_2 - \vec{v}_1\|t$ before the collision. In the case where the particles are elastic hard spheres, a collision will take place only if the impact parameter b is less than the diameter d of the spheres. The velocities after the collision are then determined by $\|\vec{v}_1 - \vec{v}_2\|, b$ and ϕ. Transformed back to the laboratory frame, the final velocities $\vec{v}_1{}', \vec{v}_2{}'$ can then be expressed as functions of $\vec{v}_1$, $\vec{v}_2$, b and ϕ.

Maxwell now assumes what the Ehrenfests later called the *Stoßzahlansatz*: the number of collisions during a time dt, say $N(\vec{v}_1, \vec{v}_2)$, in which the initial velocities $\vec{v}_1, \vec{v}_2$ within an element $d^3\vec{v}_1 d^3\vec{v}_2$ are changed into final velocities $\vec{v}_1{}', \vec{v}_2{}'$ in an element $d^3\vec{v}_1{}' d^3\vec{v}_2{}'$ within a spatial volume element $dV = bdbd\phi dz = \|\vec{v}_1 - \vec{v}_2\|\, bdbd\phi dt$ is proportional to the product of the number of particles with velocity $\vec{v}_1$ within $d^3\vec{v}_1$ (i.e. $Nf(\vec{v}_1)d\vec{v}_1$), and those with velocity $\vec{v}_2$ within $d^3\vec{v}_2$ (i.e. $Nf(\vec{v}_2)d^3\vec{v}_2$), and that spatial volume element. Thus:

$$(29)\quad N(\vec{v}_1, \vec{v}_2) = N^2 f(\vec{v}_1) f(\vec{v}_2) \|\vec{v}_2 - \vec{v}_1\| d^3\vec{v}_1 d^3\vec{v}_2 bdbd\phi dt.$$

Due to the time reversal invariance of the collision laws, a similar consideration applies to the so-called inverse collisions, in which initial velocities $\vec{v}_1{}', \vec{v}_2{}'$ and final velocities $\vec{v}_1$ and $\vec{v}_2$ are interchanged. Their number is proportional to

$$(30)\quad N(\vec{v}_1{}', \vec{v}_2{}') = N^2 f(\vec{v}_1{}') f(\vec{v}_1{}') \|\vec{v}_2{}' - \vec{v}_1{}'\| d^3\vec{v}_1{}' d\vec{v}_2{}' bdbd\phi dt$$

Maxwell argues that the distribution of velocities will remain stationary, i.e. unaltered in the course of time, if the number of collisions of these two kinds are equal, i.e. if

$$(31)\quad N(\vec{v}_1{}', \vec{v}_2{}') = N(\vec{v}_1, \vec{v}_2).$$

Moreover, the collision laws entail that $\|\vec{v}_2{}' - \vec{v}_1{}'\| = \|\vec{v}_2 - \vec{v}_1\|$ and $d^3\vec{v}_1{}' d^3\vec{v}_2{}' = d^3\vec{v}_1 d^3\vec{v}_2$. Hence, the condition (31) may be simplified to

$$(32)\quad f(\vec{v}_1) f(\vec{v}_2) = f(\vec{v}_1{}') f(\vec{v}_2{}'), \quad \text{for all } \vec{v}_1, \vec{v}_2.$$

This is the case for the Maxwellian distribution (26). Therefore, Maxwell says, the distribution (26) is a "possible" form.

He goes on to claim that it is also the *only* possible form for a stationary distribution.This claim, i.e. that stationarity of the distribution f can only arise under (32) is nowadays also called the principle of *detailed balancing* (cf. [Tolman, 1938, p. 165]).[14] Although his argument is rather brief, the idea seems to be that for a distribution violating (32), there must (because of the *Stoßzahlansatz*) be two velocity pairs[15] $\vec{v}_1, \vec{v}_2$ and $\vec{u}_1, \vec{u}_2$, satisfying $\vec{v}_1 + \vec{v}_2 = \vec{u}_1 + \vec{u}_2$ and $v_1^2 + v_2^2 = u_1^2 + u_2^2$, such that the collisions would predominantly transform $(\vec{v}_1, \vec{v}_2) \longrightarrow (\vec{u}_1, \vec{u}_2)$ rather

[14]The reader might be warned, however, that the name 'detailed balancing' is also used to cover somewhat different ideas than expressed here [Tolman, 1938, p. 521].

[15]Actually, Maxwell, discusses only velocities of a single molecule. For clarity, I have transposed his argument to a discussion of pairs.

than $(\vec{u}_1, \vec{u_2}) \longrightarrow (\vec{v}_1, \vec{v}_2)$. But then, since the distribution is stationary, there must be a third pair of velocities, $(\vec{w}_1, \vec{w}_2)$, satisfying similar relations, for which the collisions predominantly produce transitions $(\vec{u}_1, \vec{u_2}) \longrightarrow (\vec{w}_1, \vec{w}_2)$, etc. Now, the distribution can only remain stationary if any such sequence closes into a cycle. Hence there would be cycles of velocity pairs $(\vec{v}_1, \vec{v}_2) \longrightarrow (\vec{u}_1, \vec{u}_2) \longrightarrow \ldots \longrightarrow (\vec{v}_1, \vec{v}_2)$ which the colliding particles go through, eventually returning to their original velocities.

Maxwell then argues: "Now it is impossible to assign a reason why the successive velocities of a molecule should be arranged in this cycle rather than in the reverse order" [Maxwell, 1867, p.45]. Therefore, he argues, these two cycles should be equally probable, and, hence, a collision cycle of the type $(\vec{v}_1, \vec{v}_2) \longrightarrow (\vec{v}_1{}', \vec{v}_2{}')$ is already equally probable as a collision cycle of the type $(\vec{v}_1{}', \vec{v}_2{}') \longrightarrow (\vec{v}_1, \vec{v}_2)$, i.e. condition (32) holds.

Comments. First, a clear advantage of Maxwell's 1867 derivation of the distribution function (26) is that the collisions play a crucial role. The argument would not apply if there were no collisions between molecules. A second point to note is that the distribution (26) is singled out because of its *stationarity*, instead of its spherical symmetry and factorization properties. This is also a major improvement upon his previous paper, since stationarity is essential to thermal equilibrium.

A crucial element in the argument is still an assumption about independence. But now, in the *Stoßzahlansatz*, the initial velocities of colliding particles are assumed independent, instead of the orthogonal velocity components of a single particle. Maxwell does not expand on why we should assume this *ansatz*; he clearly regarded it as obvious. Yet it seems plausible to argue that he must have had in the back of his mind some version of the principle of insufficient reason, i.e., that we are entitled to treat the initial velocities of two colliding particles as independent because we have no reason to assume otherwise. Although still an argument from insufficient reason, this is at least a much more plausible application than in the 1860 paper.

A main defect of the paper is his sketchy claim that the Maxwell distribution (26) would be the unique stationary distribution. This claim may be broken in two parts: (a) the cycle argument just discussed, leading Maxwell to argue for detailed balancing; and (b) the claim that the Maxwell distribution is uniquely compatible with this condition.

A demonstration for part (b) was not provided by Maxwell at all; but this gap was soon bridged by Boltzmann (1868) — and Maxwell gave Boltzmann due credit for this proof. But part (a) is more interesting. We have seen that Maxwell here explicitly relied on reasoning from insufficient reason. He was criticized on this point by [Boltzmann, 1872] and also by [Guthrie, 1874].

Boltzmann argued that Maxwell was guilty of begging the question. If we suppose that the two cycles did not occur equally often, then this supposition by itself would provide a reason for assigning unequal probabilities to the two types of collisions.[16] This argument by Boltzmann indicates, at least in my opinion that he was much less prepared than Maxwell to argue in terms of insufficient reason. Indeed, as we shall see in Section 4, his view on probability seems much more thoroughly frequentist than Maxwell.

In fact Boltzmann later repeatedly mentioned the counterexample of a gas in which all particles are lined up so that they only collide centrally, and move perpendicularly between parallel walls [Boltzmann, 1872 (Boltzmann, 1909, I p. 358); Boltzmann, 1878 (Boltzmann, 1909, II p. 285)]. In this case, the velocity distribution

$$(33)\quad \frac{1}{2}\left(\delta(v-v_0)+\delta(v+v_0)\right)$$

is stationary too.

Some final remarks on Maxwell's work: As we have seen, it is not easy to pinpoint Maxwell's interpretation of probability. In his (1860), he identifies the probability of a particular molecular state with the relative number of particles that possess this state.[17] Yet, we have also seen that he relates probability to a state of knowledge. Thus, his position may be characterized as somewhere between the classical and the frequentist view.

Note that Maxwell never made any attempt to reproduce the second law. Rather he seems to have been content with the statistical description of thermal equilibrium in gases.[18] All his writings after 1867 indicate that he was convinced that a derivation of the Second Law from mechanical principles was impossible. Indeed, his remarks on the Second Law generally point to the view that the Second Law "has only statistical certainty" (letter to Tait, undated; [Garber *et al.*, 1995, p. 180]), and that statistical considerations were foreign to the principles of mechanics. Indeed, Maxwell was quite amused to see Boltzmann and Clausius engage in a dispute about who had been the first to reduce the Second Law of thermodynamics to mechanics:

> It is rare sport to see those learned Germans contending the priority of the discovery that the 2nd law of $\theta\Delta$cs is the 'Hamiltonsche Prinzip',

[16]More precisely, Boltzmann argued as follows: "In order to prove the impossibility [of the hypothesis] that the velocity of [a pair of] molecule[s] changes more often from $[(\vec{v}_1, \vec{v}_2)$ to $(\vec{v}_1{}', \vec{v}_2{}')]$ than the converse, Maxwell says that there should then exist a closed series of velocities that would be traversed rather in one order than the other. This, however, could not be, he claims, because one could not indicate a reason, why molecules would rather traverse the cycle in one order than the other. But it appears to me that this last claim already presupposes as proven what is to be proved. Indeed, if we assume as already proven that the velocities change as often from $(\vec{v}_1, \vec{v}_2)$ to $(\vec{v}_1{}', \vec{v}_2{}')$ as conversely, then of course there is no reason why the cycle should rather be run through in one order than the other. But if we assume that the statement to be proven is not yet proved, then the very fact that the velocities of the molecules prefer to change rather from $(\vec{v}_1, \vec{v}_2)$ to $(\vec{u}_1, \vec{u}_2)$ than conversely, rather from $(\vec{u}_1, \vec{u}_2)$ to $(\vec{w}_1, \vec{w}_2)$ than conversely, etc. would provide the reason why the cycle is traversed rather one way than the other" [Boltzmann, 1909, I, p. 319].

[17]Curiously, this terminology is completely absent in his 1867 paper.

[18]Apart from a rather lame argument in [Maxwell, 1860] analyzed by [Brush, 1976, p.344].

> [. . .] The Hamiltonsche Prinzip, the while, soars along in a region unvexed by statistical considerations, while the German Icari flap their waxen wings in nephelococcygia[19] amid those cloudy forms which the ignorance and finitude of human science have invested with the incommunable attributes of the invisible Queen of Heaven (letter to Tait, 1873; [Garber *et al.*, 1995, p. 225])

Clearly, Maxwell saw a derivation of the Second Law from pure mechanics, "unvexed by statistical considerations", as an illusion. This point appears even more vividly in his thought experiment of the "Maxwell demon", by which he showed how the laws of mechanics could be exploited to produce a violation of the Second Law. For an entry in the extensive literature on Maxwell's demon, I refer to [Earman and Norton, 1998; 1999; Leff and Rex, 2003; Bennett, 2003; Norton, 2005].

But neither did Maxwell make any effort to reproduce the Second Law on a *unified* statistical/mechanical basis. Indeed, the scanty comments he made on the topic (e.g. in [Maxwell, 1873; Maxwell, 1878b]) rather seem to point in another direction. He distinguishes between what he calls the 'statistical method' and the 'historical' or 'dynamical' (or sometimes 'kinetic') method. These are two modes of description for the same system. But rather than unifying them, Maxwell suggests they are competing, or even incompatible — one is tempted to say "complementary"– methods, and that it depends on our own knowledge, abilities, and interests which of the two is appropriate. For example:

> In dealing with masses of matter, while we do not perceive the individual molecules, we are *compelled* to adopt what I have described as the statistical method, and to abandon the strict dynamical method, in which we follow every motion by the calculus [Maxwell, 1872, p. 309, emphasis added].

In this respect, his position stands in sharp contrast to that of Boltzmann, who made the project of finding this unified basis his lifework.

4 BOLTZMANN[20]

4.1 Early work: Stoßzahlansatz *and ergodic hypothesis*

Boltzmann had already been considering the problem of finding a mechanical derivation of the Second Law in a paper of 1866. At that time, he did not know of Maxwell's work. But in 1868, he had read both Maxwell's papers of 1860 and 1867. Like Maxwell, he focuses on the study of gases in thermal equilibrium, instead of the Second Law. He also adopts Maxwell's idea of characterizing thermal equilibrium by a probability distribution, and the *Stoßzahlansatz* as the central

[19]'Cloudcuckooland", an illusory place in Aristophanes' *The Birds*.

[20]Parts of this section were also published in [Uffink, 2004].

dynamical assumption. But along the way in this extensive paper, Boltzmann comes to introduce an entirely different alternative approach, relying on what we now call the ergodic hypothesis.

As we saw in section 3.3, Maxwell had derived his equilibrium distribution for two special gas models (i.e. a hard sphere gas in 1860 and a model of point particles with a central r^5 repulsive force acting between them in 1867). He had noticed that the distribution, once attained, will remain stationary in time (when the gas remains isolated), and also argued (but not very convincingly) that it was the *only* such stationary distribution.

In the first section of his 1868a, Boltzmann aims to reproduce and improve these results for a system of an infinite number of hard discs moving in a plane. He regards it as obvious that the equilibrium distribution should be independent of the position of the discs, and that every direction of their velocities is equally probable. It is therefore sufficient to consider the probability distribution over the various values of the velocity $v = \|\vec{v}\|$. However, Boltzmann started out with a somewhat different interpretation of probability in mind than Maxwell. He introduced the probability distribution as follows:

> Let $\phi(v)dv$ be the sum of all the instants of time during which the velocity of a disc in the course of a very long time lies between v and $v + dv$, and let N be the number of discs which on average are located in a unit surface area, then
>
> (34) $N\phi(v)dv$
>
> is the number of discs per unit surface whose velocities lie between v and $v + dv$ [Boltzmann, 1909, I, p. 50].[21]

Thus, $\phi(v)dv$ is introduced as the relative *time* during which a (given) disc has a particular velocity. But, in the same breath, this is identified with the relative *number* of discs with this velocity.

This remarkable quote shows how he identified two different meanings for the same function. We shall see that this equivocation returned in different guises again and again in Boltzmann's writings.[22] Indeed, it is, I believe, the very heart of the ergodic problem, put forward so prominently by the Ehrenfests (cf. paragraph 6.1). Either way, of course, whether we average over time or particles, probabilities are defined here in strictly mechanical terms, and are therefore objective properties of the gas.

[21] Here and below, "Abh." refers to the three volumes of Boltzmann's collected scientific papers [Boltzmann, 1909].

[22] This is not to say that he always conflated these two interpretations of probability. Some papers employ a clear and consistent choice for one interpretation only. But then that choice differs between papers, or even in different sections of a single paper. In fact, in [Boltzmann, 1871c] he even multiplied probabilities with different interpretations into one equation to obtain a joint probability. But then in 1872 he conflates them again. Even in his last paper [Boltzmann and Nabl, 1904], Boltzmann identifies two meanings of probability with a simple-minded argument.

Next he goes into a detailed mechanical description of a two-disc collision process. If the variables which specify the initial velocities of two discs before the collision lie within a given infinitesimal range, Boltzmann determines how the collision will transform the initial values of these variables $(\vec{v}_i, \vec{v}_j)$ into the final values $(\vec{v}'_i, \vec{v}'_j)$ in another range. At this point a two-dimensional analogy of the *Stoßzahlansatz* is introduced to obtain the number of collisions per unit of time. As in Maxwell's treatment, this amounts to assuming that the number of such collisions is proportional to the product $\phi(v_1)\phi(v_2)$. In fact:

$$(35)\quad N(\vec{v}_1, \vec{v}_2) \propto N^2 \frac{\phi(v_1)}{v_1}\frac{\phi(v_2)}{v_2}\|\vec{v}_2 - \vec{v}_1\| dv_1 dv_2 dt$$

where the proportionality constant depends on the geometry of the collision.

He observes that if, for all velocities v_i, v_j and all pairs of discs i, j, the collisions that transform the values of the velocities (v_i, v_j) from a first range $dv_i dv_j$ into values v'_i, v'_j within the range $dv'_i dv'_j$ occur equally often as conversely (i.e., equally often as those collisions that transform initial velocities v'_i, v'_j within $dv'_i dv'_j$ into final values v_i, v_j within $dv_i dv_j$), the distribution ϕ will remain stationary. He states "This distribution is therefore the desired one" [Boltzmann, 1909, I p. 55]. Actually, this is the first occasion in the paper at which the desideratum of stationarity of the probability distribution is mentioned.

Using the two-dimensional version of the *Stoßzahlansatz* this desideratum leads to

$$(36)\quad \frac{\phi(v_i)}{v_i}\frac{\phi(v_j)}{v_j} = \frac{\phi(v'_i)}{v'_i}\frac{\phi(v'_j)}{v'_j}$$

He shows [Boltzmann, 1909, p. 57] that the only function obeying condition (36) for all choices of $v_1, v_2, v'_1 v'_2$, compatible with the energy equation $v_1^2 + v_2^2 = v_1'^2 + v_2'^2$, is of the form

$$(37)\quad \phi(v) = 2hve^{-hv^2},$$

for some constant h. Putting $f(v) := v\phi(v)$ we thus obtain the two-dimensional version of the Maxwell distribution (26). Boltzmann does not address the issue of whether the condition (36) is necessary for the stationarity of ϕ.

In the next subsections of 1868a, Boltzmann repeats the derivation, each time in slightly different settings. First, he goes over to the three-dimensional version of the problem, assuming a system of hard spheres, and supposes that one special sphere is accelerated by an external potential $V(\vec{x})$. He shows that if the velocities of all other spheres are distributed according to the Maxwellian distribution (26), the probability distribution of finding the special sphere at place $\vec{x}$ and velocity $\vec{v}$ is $f(\vec{v}, \vec{x}) \propto e^{-h(\frac{1}{2}mv^2 + V(\vec{x}))}$ [Boltzmann, 1909, I, p.63]. In a subsequent subsection, he replaces the spheres by material points with a short-range interaction potential and reaches a similar result.

At this point, (the end of Section I of the [1868a] paper), the argument suddenly switches course. Instead of continuing in this fashion, Boltzmann announces

[Boltzmann, 1909, p. 80] that all the cases treated, and others yet untreated, follow from a much more general theorem. This theorem, which, as we shall see relies on the ergodic hypothesis, is the subject of the second and third Section of the paper. I will limit the discussion to the third section and rely partly on Maxwell's (1879) exposition, which is somewhat simpler and clearer than Boltzmann's own.

The ergodic hypothesis

Consider a general mechanical system of N material points, each with mass m, subject to an arbitrary time-independent potential.[23] In modern notation, let $x = (\vec{q}_1, \vec{p}_1; \ldots; \vec{q}_N, \vec{p}_N)$ denote the canonical position coordinates and momenta of the system. Its Hamiltonian is then[24]

$$(38)\quad H(x) = \frac{1}{2m}\sum_i^N \vec{p}_i^{\,2} + U(\vec{q}_1, \ldots, \vec{q}_N).$$

The state x may be represented as a phase point in the mechanical phase space Γ. Under the Hamiltonian equations of motion, this phase point evolves in time, and thus describes a trajectory x_t ($t \in \mathbb{R}$). This trajectory is constrained to lie on a given energy hypersurface $\Gamma_E = \{x \in \Gamma \ : \ H(x) = E\}$. Boltzmann asks for the probability (i.e. the fraction of time during a very long period) that the phase point lies in a region $dx = d^3\vec{q}_1 \cdots d^3\vec{p}_N$, which we may write as:

$$(39)\quad \rho(x)dx = \chi(x)\delta(H(x) - E)dx.$$

for some function χ. Boltzmann seems to assume implicitly that this distribution is stationary. This property would of course be guaranteed if the "very long period" were understood as an infinite time. He argues, by Liouville's theorem, that χ is a constant for all points on the energy hypersurface that are "possible", i.e. that are actually traversed by the trajectory. For all other points χ vanishes. If we neglect those latter points, the function χ must be constant over the entire energy hypersurface, and the probability density ρ takes the form

$$(40)\quad \rho_{\mathrm{mc}}(x) = \frac{1}{\omega(E)}\delta(H(x) - E),$$

the micro-canonical distribution, where

$$(41)\quad \omega(E) = \int \delta H(x) = E dx$$

is the so-called structure function.

In particular, one can now form the marginal probability density for the positions $\vec{q}_1, \ldots, \vec{q}_N$ by integrating over the momenta:

[23] Although Boltzmann does not mention it at this stage, his previous section added the stipulation that the particles are enclosed in a finite space, surrounded by perfectly elastic walls.

[24] Actually Boltzmann allows the N masses to be different but restricts the potential as being due to external and mutual two-particle forces only, i.e. $H(x) = \sum_i \frac{\vec{p}_i^{\,2}}{2m_i} + \sum_{i \leq j} U_{ij}(\|\vec{q}_i - \vec{q}_j\|) + \sum_i U_i(\vec{q}_i)$.

$$(42)\quad \rho_{\rm mc}(\vec{q}_1,\ldots,\vec{q}_N) := \int \rho_{\rm mc}(x)\, d^3\vec{p}_1\cdots d^3\vec{p}_N = \frac{2m}{\omega(E)}\int \delta\left(\sum \vec{p}_i^{\,2} - 2m(E-U(q))\right) d\vec{p}_1\cdots d\vec{p}_N.$$

The integral over the momenta can be evaluated explicitly (it is $2R^{-1}$ times the surface area of a hypersphere with radius $R=\sqrt{2m(E-U)}$ in $n=3N$ dimensions), to obtain

$$(43)\quad \rho_{\rm mc}(\vec{q}_1,\ldots,\vec{q}_N) = \frac{2m(\pi)^{n/2}}{\omega(E)\Gamma(\frac{n}{2})}(2m(E-U(q))^{(n-2)/2},$$

where Γ denotes Euler's gamma function: $\Gamma(x) := \int_0^\infty t^{x-1}e^{-t}dt$.

Similarly, the marginal probability density for finding the first particle with a given momentum component p_{1x} as well as finding the positions of all particles at $\vec{q}_1,\ldots,\vec{q}_N$ is

$$\begin{aligned}\rho_{\rm mc}(p_{1x},\vec{q}_1\ldots,\vec{q}_N) &= \int \rho_{\rm mc}(x)\, dp_{1y}dp_{1z}d^3\vec{p}_2\cdots d^3\vec{p}_N \\ (44)\qquad &= \frac{2m\pi^{(n-1)/2}}{\omega(E)\Gamma(\frac{n-1}{2})}\left(2m(E-U(q)) - p_{1x}^2\right)^{(n-3)/2}.\end{aligned}$$

These two results can be conveniently presented in the form of the conditional probability that the x-component of momentum of the first particle has a value between p and $p+dp$, given that the positions have the values $\vec{q}_1\ldots,\vec{q}_N$, by taking the ratio of (44) and (43):

$$(45)\quad \rho_{\rm mc}(p\,|\,\vec{q}_1,\ldots,\vec{q}_N)dp = \frac{1}{\sqrt{2m\pi}}\frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})}\frac{(E-U-\frac{p^2}{2m})^{(n-2)/2}}{(E-U)^{(n-3)/2}}dp.$$

This, in essence, is the general theorem Boltzmann had announced. Further, he shows that in the limit where $n\longrightarrow\infty$, and the kinetic energy per degree of freedom $\kappa := (E-U)/n$ remains constant, the expression (45) approaches

$$(46)\quad \frac{1}{\sqrt{4\pi m\kappa}}\exp\left(-\frac{p^2}{4m\kappa}\right)dp.$$

This probability thus takes the same form as the Maxwell distribution (26), if one equates $\kappa = \frac{1}{2}kT$. Presumably, it is this result that Boltzmann had in mind when he claimed that all the special cases he has discussed in section 1 of his paper, would follow from the general theorem. One ought to note however, that since U, and therefore κ, depends on the coordinates, the condition κ = constant is different for different values of $(\vec{q}_1,\ldots,\vec{q}_n)$.

Some comments on this result.

1. The difference between this approach and that relying on the *Stoßzahlansatz* is rather striking. Instead of concentrating on a gas model in which particles

are assumed to move freely except for their occasional collisions, Boltzmann here assumes a much more general Hamiltonian model with an arbitrary interaction potential $U(\vec{q}_1, \ldots \vec{q}_N)$. Moreover, the probability density ρ is defined over phase space, instead of the space of molecular velocities. This is the first occasion where probability considerations are applied to the state of the mechanical system as whole, instead of its individual particles. If the transition between kinetic gas theory and statistical mechanics may be identified with this caesura, (as argued by the Ehrenfests 1912 and by Klein 1973) it would seem that the transition has already been made right here.

But of course, for Boltzmann the transition did not involve a major conceptual move, thanks to his conception of probability as a relative time. Thus, the probability of a particular state of the total system is still identified with the fraction of time in which that state is occupied by the system. In other words, he had no need for ensembles or non-mechanical probabilistic assumptions.

However, one should note that the equivocation between relative time and relative number of particles, which was comparatively harmless in the first section of the 1868 paper, is now no longer possible in the interpretation of ρ. Consequently, the conditional probability $\rho(p|\vec{q}_1, \ldots \vec{q}_N)$ gives us the relative time that the total system is in a state for which particle 1 has a momentum with x-component between p and $p + dp$, for given values of all the positions. There is no immediate route to conclude that this has anything to do with the relative number of particles with the momentum p. In fact, there is no guarantee that the probability (45) for particle 1 will be the same for other particles too, unless we use the assumption that U is invariant under permutation of the particles. Thus, in spite of their identical form, the distribution (46) has a very different meaning than (26).

2. The transition from (45) to (46), by letting the number of particles become infinite, also seems to be the first instance of a thermodynamical limit. Since the Maxwell distribution is thus recovered only in this limit, Boltzmann's procedure resolves some questions raised above concerning Maxwell's distribution. For a finite number of particles, the distribution (45) always has a finite support, i.e. $\rho_{mc} = 0$ for those values of $p_i^2 \geq 2m(E - U)$. Thus, we do not run into trouble with the finite amount of energy in the gas.

3. Most importantly, the results (45,46) open up a perspective of great generality. It suggests that the probability of the molecular velocities for an isolated system in a stationary state will always assume the Maxwellian form if the number of particles tends to infinity. Notably, this proof seems to completely dispense with any particular assumption about collisions, or other details of the mechanical model involved, apart from the assumption that it is Hamiltonian. Indeed it need not even represent a gas.

4. The main weakness of the present result is its assumption that the trajectory actually visits all points on the energy hypersurface. This is nowadays called the *ergodic hypothesis*.[25]

[25] The literature contains some surprising confusion about how the hypothesis got its name. The Ehrenfests borrowed the name from Boltzmann's concept of an "*Ergode*", which he introduced

Boltzmann returned to this issue on the final page of the paper [Boltzmann, 1909, p. 96]. He notes there that there might be exceptions to his theorem, for example, when the trajectory is periodic. However, Boltzmann observed, such cases would be sensitive to the slightest disturbance from outside. They would be destroyed, e.g. by the interaction of a single free atom that happened to be passing by. He argued that these exceptions would thus only provide cases of unstable equilibrium.

Still, Boltzmann must have felt unsatisfied with his own argument. According to an editorial footnote in his collected works [Boltzmann, 1909, I p.96], Boltzmann's personal copy of the paper contains a hand-written remark in the margin stating that the point was still dubious and that it had not been proven that, even in the presence of interaction with a single external atom, the system would traverse all possible values compatible with the energy equation.

Doubts about the ergodic hypothesis

Boltzmann's next paper 1868b was devoted to checking the validity of the ergodic hypothesis in a relatively simple solvable mechanical model. This paper also gives a nice metaphoric formulation of the ergodic hypothesis: if the phase point were a light source, and its motion exceedingly swift, the entire energy surface would appear to us as homogeneously illuminated [Boltzmann, 1909, I, p. 103]. However, his doubts were still not laid to rest. His next paper on gas theory 1871a returns to the study of a detailed mechanical gas model, this time consisting of polyatomic

---

in [Boltzmann, 1884] and also discussed in his Lectures on Gas Theory [Boltzmann, 1898]. But what did Boltzmann actually understood by an *Ergode*? Brush points out in his translation of [Boltzmann, 1898, p. 297], and similarly in [Brush, 1976, p. 364], that Boltzmann used the name to denote a stationary ensemble, characterized by the microcanonical distribution in phase space. In other words, in in the context of Boltzmann's 1898 an *Ergode* is just an microcanonical ensemble, and seems to have nothing to do to do with the so-called ergodic hypothesis. Brush criticized the Ehrenfests for causing confusion by their terminology.

However, in his original 1884 introduction of the phrase, the name *Ergode* is used for a stationary ensemble with only a *single* integral of motion , i.e. its total energy. As a consequence, the ensemble is indeed micro-canonical, but, what is more, every member of the ensemble satisfies the hypothesis of traversing every phase point with the given total energy. Indeed, in this context, being an element of an *Ergode* implies satisfaction of this hypothesis. Thus, the Ehrenfests were actually justified in baptizing the hypothesis "ergodic".

Another dispute has emerged concerning the etymology of the term. The common opinion, going back at least to the Ehrenfests has been that the word derived from *ergos* (work) and *hodos* (path). [Gallavotti, 1994] has argued however that "undoubtedly" it derives from *ergos* and *eidos* (similar). Now one must grant Gallavotti that it ought to expected that the etymology of the suffix "-ode" of ergode is identical to that of other words Boltzmann coined in this paper, like *Holode, Monode, Orthode* and *Planode*; and that a reference to path would be somewhat unnatural in these last four cases. However, I don't believe a reference to "eidos" would be more natural. Moreover, it seems to me that if Boltzmann intended this etymology, he would have written "Ergoide" in analogy to planetoid, ellipsoid etc. That he was familiar with this common usage is substantiated by him coining the term "*Momentoide*" for momentum-like degrees of freedom (i.e. those that contribute a quadratic term to the Hamiltonian) in [Boltzmann, 1892]. The argument mentioned by [Cercignani, 1998, p. 141] (that Gallavotti's father is a classicist) fails to convince me in this matter.

molecules, and avoids any reliance on the ergodic hypothesis. And when he did return to the ergodic hypothesis in 1871b, it was with much more caution. Indeed, it is here that he actually first described the worrying assumption as an *hypothesis*, formulated as follows:

> The great irregularity of the thermal motion and the multitude of forces that act on a body make it probable that its atoms, due to the motion we call heat, traverse all positions and velocities which are compatible with the principle of [conservation of] energy [Boltzmann, 1909, I p. 284].[26]

Note that Boltzmann formulates this hypothesis for an arbitrary body, i.e. it is not restricted to gases. He also remarks, at the end of the paper, that "the proof that this hypothesis is fulfilled for thermal bodies, or even is fullfillable, has not been provided" [Boltzmann, I p. 287].

There is a major confusion among modern commentators about the role and status of the ergodic hypothesis in Boltzmann's thinking. Indeed, the question has often been raised how Boltzmann could ever have believed that a trajectory traverses *all* points on the energy hypersurface, since, as the Ehrenfests conjectured in 1911, and was shown almost immediately in 1913 by Plancherel and Rosenthal, this is mathematically impossible when the energy hypersurface has a dimension larger than 1 (cf. paragraph 6.1).

It is a fact that both his (1868a, Abh. I, p.96) and (1871b, Abh. I, p.284) mention external disturbances as an ingredient in the motivation for the ergodic hypothesis. This might be taken as evidence for 'interventionism', i.e. the viewpoint that such external influences are crucial in the explanation of thermal phenomena (cf: [Blatt, 1959; Ridderbos and Redhead, 1998]). Yet even though Boltzmann clearly expressed the thought that these disturbances might help to motivate the ergodic hypothesis, he never took the idea very seriously. The marginal note in the (1868a) paper mentioned above indicated that, even if the system is disturbed, there is still no easy proof of the ergodic hypothesis, and all his further investigations concerning this hypothesis assume a system that is either completely isolated from its environment, or at most acted upon by a static external force. Thus, interventionism did not play a significant role in his thinking.[27]

It has also been suggested, in view of Boltzmann's later habit of discretizing continuous variables, that he somehow thought of the energy hypersurface as a discrete manifold containing only finitely many discrete cells [Gallavotti, 1994]. On this reading, obviously, the mathematical no-go theorems of Rosenthal and Plancherel no longer apply. Now it is definitely true that Boltzmann developed a preference towards discretizing continuous variables, and would later apply this

[26]An equivalent formulation of the ergodic hypothesis is that the Hamiltonian is the only independent integral of the Hamiltonian equations of motion. This version is given in the same paper [Boltzmann, 1909, p. 281-2]

[27]Indeed, on the rare occasions on which he later mentioned external disturbances, it was only to say that they are "not necessary" [Boltzmann, 1895b]. See also [Boltzmann, 1896, §91].

procedure more and more (although usually adding that this was fictitious and purely for purposes of illustration and more easy understanding, cf. paragraph 4.2). However, there is no evidence in the (1868) and (1871b) papers that Boltzmann implicitly assumed a discrete structure of the mechanical phase space or the energy hypersurface.

Instead, the context of his 1871b makes clear enough how he intended the hypothesis, as has already been argued by [Brush, 1976]. Immediately preceding the section in which the hypothesis is introduced, Boltzmann discusses trajectories for a simple example: a two-dimensional harmonic oscillator with potential $U(x,y) = ax^2 + by^2$. For this system, the configuration point (x, y) moves through the surface of a rectangle. (Cf. Fig. 2. See also [Cercignani, 1998, p. 148].) He then notes that if a/b is rational, (actually: if $\sqrt{a/b}$ is rational) this motion is periodic. However, if this value is irrational, the trajectory will, in the course of time, traverse "*almählich die ganze Fläche*" [Boltzmann, 1909, p. 271] of the rectangle. He says that in this case x and y are *independent*, since for each values of x an infinity of values for y in any interval in its range are possible. The very fact that Boltzmann considers intervals for the values of x and y of arbitrary small sizes, and stressed the distinction between rational and irrational values of the ratio a/b, indicates that he did *not* silently presuppose that phase space was essentially discrete, where those distinctions would make no sense.

Now clearly, in modern language, one should say that if $\sqrt{a/b}$ is irrational the trajectory is *dense* in the rectangle, but not that it traverses all points. Boltzmann did not possess this language. In fact, he could not have been aware of Cantor's insight that the continuum contains more than a countable infinity of points. Thus, the correct statement that, in the case that $\sqrt{a/b}$ is irrational, the trajectory will traverse, for each value of x, an infinity of values of y within any interval however small, could easily have led him to believe (incorrectly) that *all* values of x and y are traversed in the course of time.

It thus seems eminently plausible, in view of the fact that this discussion immediately precedes the formulation of the ergodic hypothesis, that Boltzmann's understanding of the ergodic hypothesis is really what Ehrenfests dubbed the *quasi-ergodic hypothesis*: the assumption that the trajectory is dense (i.e. passes arbitrarily close to every point) on the energy hypersurface.[28] The quasi-ergodic hypothesis is not mathematically impossible in higher-dimensional phase spaces. However, the quasi-ergodic hypothesis does not entail the desired conclusion that the only stationary probability distribution over the energy surface is micro-canonical.

Nevertheless, Boltzmann remained sceptical about the validity of his hypothesis, and attempted to explore different routes to his goal of characterizing thermal equilibrium in mechanics. Indeed, both the preceding 1871a and his next paper 1871c present alternative arguments, with the explicit recommendation that they avoid hypotheses. In fact, he did not return to the ergodic hypothesis at all until

[28]Or some hypothesis compatible with the quasi-ergodic hypothesis. As it happens, Boltzmann's example is also compatible with the measure-theoretical hypothesis of 'metric transitivity' (cf. paragraph 6.1).

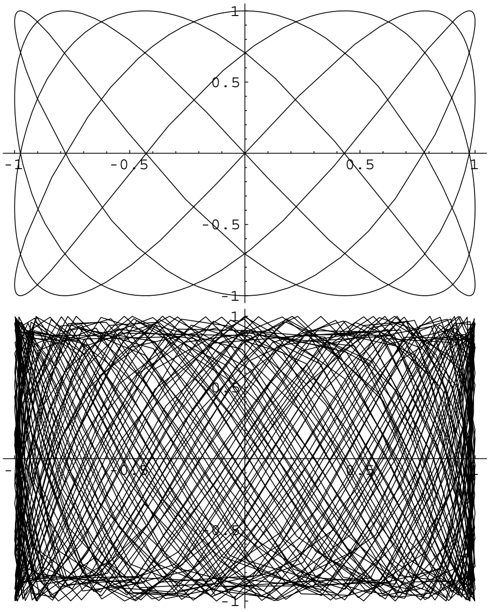


Figure 2. Trajectories in configuration space for a two-dimensional harmonic oscillator with potential $U(x, y) = ax^2 + by^2$. Illustrating the distinction between (i) the case where $\sqrt{a/b}$ is rational (here 4/7) and (ii) irrational ($1/e$). Only a fragment of the latter trajectory has been drawn.

the 1880s (stimulated by Maxwell's 1879 review of the last section of Boltzmann's 1868 paper). At that time, perhaps feeling fortified by Maxwell's authority, he was to express much more confidence in the ergodic hypothesis. However, after 1885, this confidence disappears again, and although he mentions the hypothesis occasionally in later papers, he never assumes its validity. Most notably, the ergodic hypothesis is not even mentioned in his *Lectures on Gas Theory* 1896, 1898.

To sum up, what role did the ergodic hypothesis play for Boltzmann? It seems that Boltzmann regarded the ergodic hypothesis as a special dynamical assumption that may or may not be true, depending on the nature of the system, and perhaps also on its initial state and the disturbances from its environment. Its role was simply to help derive a result of great generality: for any system for which the hypothesis is true, its equilibrium state is characterized by (45), from which an analogy to the Maxwell distribution may be recovered in the limit $N \longrightarrow \infty$, regardless of any details of the inter-particle interactions, or indeed whether the system represented is a gas, fluid, solid or any other thermal body.

As we discussed in paragraph 1.4, the Ehrenfests 1912 have suggested that the ergodic hypothesis played a much more fundamental role. In particular, if the hypothesis is true, averaging over an (infinitely) long time would be identical to phase averaging with the microcanonical distribution. Thus, they suggested that Boltzmann relied on the ergodic hypothesis in order to equate time averages and phase averages, or in other words, to equate two meanings of probability (relative time and relative volume in phase space.) There is however *no* evidence that Boltzmann ever followed this line of reasoning neither in the 1870s, nor later. He simply never gave any justification for equivocating time and particle averages, or phase averages, at all. Presumably, he thought nothing much depended on this issue and that it was a matter of taste.

4.2 The Boltzmann equation and H-theorem (1872)

In 1872 Boltzmann published one of his most important papers. It contained two celebrated results nowadays known as the Boltzmann equation and the H-theorem. The latter result was the basis of Boltzmann's renewed claim to have obtained a general theorem corresponding to the Second Law. This paper has been studied and commented upon by numerous authors, and an entire translation of the text has been provided by [Brush, 1966]. Thus, for the present purposes, a succinct summary of the main points might have been sufficient. However, there is still dispute among modern commentators about its actual content.

The issue at stake in this dispute is the question whether the results obtained in this paper are presented as necessary consequences of the mechanical equations of motion, or whether Boltzmann explicitly acknowledged that they would allow for exceptions. Klein has written:

> I can find no indication in his 1872 memoir that Boltzmann conceived of possible exceptions to the H-theorem, as he later called it [Klein, 1973, p. 73].

Klein argues that Boltzmann only came to acknowledge the existence of such exceptions thanks to Loschmidt's critique in 1877. An opposite opinion is expressed by [von Plato, 1994]. Calling Klein's view a "popular image", he argues that, already in 1872, Boltzmann was well aware that his H-theorem had exceptions, and thus "already had a full hand against his future critics". Indeed, von Plato states that

> Contrary to a widely held opinion, Boltzmann is not in 1872 claiming that the Second Law and the Maxwellian distribution are *necessary* consequences of kinetic theory [von Plato, 1994, p. 81].

So it might be of some interest to try and settle this dispute.

Boltzmann (1872) starts with an appraisal of the role of probability theory in the context of gas theory. The number of particles in a gas is so enormous, and their movements are so swift that we can observe nothing but average values. The determination of averages is the province of probability calculus. Therefore, "the problems of the mechanical theory of heat are really problems in probability calculus” [Boltzmann, 1909, I, p. 317]. But, Boltzmann says, it would be a mistake to believe that the theory of heat would therefore contain uncertainties.

He emphasizes that one should not confuse incompletely proven assertions with rigorously derived theorems of probability theory. The latter are necessary consequences of their premisses, just like in any other theory. They will be confirmed by experience as soon as one has observed a sufficiently large number of cases. This last condition, however, should be no significant problem in the theory of heat because of the enormous number of molecules in macroscopic bodies. Yet, in this context, one has to make doubly sure that we proceed with the utmost rigour.

Thus, the message expressed in the opening pages of this paper seems clear enough: the results Boltzmann is about to derive are advertised as doubly checked and utterly rigorous. Still, they are theoretical. Their relationship with experience might be less secure, since any probability statement is only reproduced in observations by sufficiently large numbers of independent data. Thus, Boltzmann would have allowed for exceptions in the relationship between theory and observation, but not in the relation between premisses and conclusion.

He continues by saying what he means by probability, and repeats its equivocation as a fraction of time and the relative number of particles that we have seen earlier in 1868:

> If one wants [...] to build up an exact theory [...] it is before all necessary to determine the probabilities of the various states that one and the same molecule assumes in the course of a very long time, and that occur simultaneously for different molecules. That is, one must calculate how the number of those molecules whose states lie between certain limits relates to the total number of molecules [Boltzmann, 1909, I p. 317].

However, this equivocation is not vicious. For most of the paper the intended meaning of probability is always the relative number of molecules with a particular molecular state. Only at the final stages of his paper [Boltzmann, 1909, I, p. 400] does the time-average interpretation of probability (suddenly) recur.

Boltzmann says that both Maxwell and he had attempted the determination of these probabilities for a gas system but without reaching a complete solution. Yet, on a closer inspection, “it seems not so unlikely that these probabilities can be derived on the basis of the equations of motion alone...” [Boltzmann, 1909, I, p. 317]. Indeed, he announces, he has solved this problem for gases whose molecules consist of an arbitrary number of atoms. His aim is to prove that whatever the initial distribution of state in such a system of gas molecules, it must inevitably approach the distribution characterized by the Maxwellian form (ibid. p. 320).

The next section specializes to the simplest case of monatomic gases and also provides a more complete specification of the problem he aims to solve. The gas molecules are contained in a fixed vessel with perfectly elastic walls. They interact with each other only when they approach each other at very small distances. These interactions can be mimicked as collisions between elastic bodies. Indeed, these bodies are modeled as hard spheres [Boltzmann, 1909, I, p. 320]. Boltzmann represents the state of the gas by a time-dependent distribution function $f_t(\vec{v})$, called the "distribution of state", which gives us, at each time t, the relative number of molecules with velocity between $\vec{v}$ and $\vec{v} + d^3\vec{v}$.[29]

He also states two more special assumptions:

1. Already in the initial state of the gas, each direction of velocity is equally probable. That is:

   (47) $f_0(\vec{v}) = f_0(v)$.

   It is assumed as obvious that this will also hold for any later time.

2. The gas is spatially uniform within the container. That is, the relative number of molecules with their velocities in any given interval, and their positions in a particular spatial region R does not depend on the location of R in the available volume.

The next and crucial assumption used by Boltzmann to calculate the change in the number of particles with a velocity $\vec{v}_1$ per unit time, is the *Stoßzahlansatz*, (29) and (30).

For modern readers, there are also a few unstated assumptions that go into the construction of this equation. First, the number of molecules must be large enough so that the (discrete) distribution of their velocities can be well approximated by a continuous and differentiable function f. Secondly, f changes under the effect of binary collisions only. This means that the density of the gas should be low (so that three-particle collisions can be ignored) but not too low (which would make collisions too infrequent to change f at all). These two requirements are already hard enough to put in a mathematically precise form. The modern explicitation is that of taking the so-called Boltzmann-Grad limit (cf. paragraph 6.4). The final (unstated) assumption is that all the above assumptions remain valid in the course of time.

He addresses his aim by constructing a differentio-integral evolution equation for f_t, by taking the difference of (29) and (30) and integrating over all variables except $\vec{v}_1$ and t. The result (in a modern notation) is the *Boltzmann equation*:

$$(48)\quad \frac{\partial f_t(\vec{v}_1)}{\partial t} = N \int_0^d b db \int_0^{2\pi} d\phi \int_{\mathbb{R}^3} d^3\vec{v}_2 \, \|\vec{v}_2 - \vec{v}_1\| \left(f_t(\vec{v}_1{}') f_t(\vec{v}_2{}') - f_t(\vec{v}_1) f_t(\vec{v}_2) \right)$$

which describes the change of f in the course of time, when this function at some initial time is given. (Recall from paragraph 3.3 that the primed velocities are to be thought of as functions of the unprimed velocities and the geometrical parameters of the collision: $\vec{v}_i' = \vec{v}_i{}'(\vec{v}_1, \vec{v}_2, b, \phi)$, and d denotes the diameter of the hard spheres.)

[29] Actually Boltzmann formulated the discussion in terms of a distribution function over kinetic energy rather than velocity. I have transposed this into the latter, nowadays more common formulation.

The H-theorem

Assuming that the Boltzmann equation (48) is valid for all times, one can prove, after a few well-known manipulations, that the following quantity

(49) $H[f_t] := \int f_t(\vec{v}) \ln f_t(\vec{v}) d^3\vec{v}$

decreases monotonically in time, i.e.

(50) $\frac{dH[f_t]}{dt} \leq 0;$

as well as its stationarity for the Maxwell distribution, i.e.:

(51) $\frac{dH[f_t]}{dt} = 0 \quad (\forall t) \text{ iff } f_t(v) = Ae^{-Bv^2}.$

Boltzmann concludes Section I of the paper as follows:

> It has thus been rigorously proved that whatever may have been the initial distribution of kinetic energy, in the course of time it must necessarily approach the form found by Maxwell. [...] This [proof] actually gains much in significance because of its applicability to the theory of multi-atomic gas molecules. There too, one can prove for a certain quantity [H] that, because of the molecular motion, this quantity can only decrease or in the limiting case remain constant. Thus, one may prove that because of the atomic movement in systems consisting of arbitrarily many material points, there always exists a quantity which, due to these atomic movements, cannot increase, and this quantity agrees, up to a constant factor, exactly with the value that I found in [[Boltzmann, 1871c]] for the well-known integral $\int dQ/T$.
>
> This provides an analytical proof of the Second Law in a way completely different from those attempted so far. Up till now, one has attempted to proof that $\int dQ/T = 0$ for a reversible (*umkehrbaren*) cyclic[30] process, which however does not prove that for an irreversible cyclic process, which is the only one that occurs in nature, it is always negative; the reversible process being merely an idealization, which can be approached more or less but never perfectly. Here, however, we immediately reach the result that $\int dQ/T$ is in general negative and zero only in a limit case... [Boltzmann, 1909, I, p. 345]

Thus, as in his 1866 paper, Boltzmann claims to have a rigorous, analytical and general proof of the Second Law. From our study of the paper until now, (i.e. section I) it appears that Klein's interpretation is more plausible than von Plato's. I postpone a further discussion of this dispute to paragraph 4.2, after a brief look at the other sections of the paper.

[30]The term "cyclic" is missing in Brush's translation, although the original text does speak of "Kreisprozeß". The special notation $\oint$ for cyclic integrals was not introduced until much later.

Further sections of Boltzmann (1872)

Section II is entitled "Replacement of integrals by sums" and devoted to a repetition of the earlier arguments, now assuming that the kinetic energies of the molecules can only take values in a discrete set $\{0, \epsilon, 2\epsilon, \ldots, p\epsilon\}$. Boltzmann shows that in the limit $\epsilon \longrightarrow 0$, $p\epsilon \longrightarrow \infty$ the same results are recovered.

Many readers have been surprised by this exercise, which seems rather superfluous both from a didactic and a logical point of view. (However, some have felt that it foreshadowed the advent of quantum theory.) Boltzmann offers as motivation for the detour that the discrete approach is clearer than the previous one. He argues that integrals only have a symbolic meaning, as a sum of infinitely many infinitesimal elements, and that a discrete calculation yields more understanding. He does not argue, however, that it is closer to physical reality. Be that as it may, the section does eventually take the limit, and recovers the same results as before.

The third section treats the case where the gas is non-uniform, i.e., when condition 2 above is dropped. For this case, Boltzmann introduces a generalized distribution function $f_t(\vec{r}, \vec{v})$, such that $f_t d^3\vec{r} d^3\vec{v}$ represents the relative number of particles with a position in a volume element $d^3\vec{r}$ around $\vec{r}$ and a velocity in an element $d^3\vec{v}$ around $\vec{v}$.

He obtains a corresponding generalized Boltzmann equation:

$$\frac{\partial f_t(\vec{r}, \vec{v})}{\partial t} + \vec{v} \cdot \nabla_x f_t + \frac{\vec{F}}{m} \cdot \nabla_v f_t =$$

$$N \int b db d\phi d^3\vec{v}_2 \, \|\vec{v}_2 - \vec{v}_1\| \Big(f_t(\vec{r}, \vec{v}_1')) f_t(\vec{r}, \vec{v}_2') - f_t(\vec{r}, \vec{v}_1)) f_t(\vec{r}, \vec{v}_2) \Big) \tag{52}$$

where $\vec{F}$ denotes an external force field on the gas. The quantity H now takes the form $H[f_t] := \int f_t(\vec{r}, \vec{v}) d^3\vec{r} d^3\vec{v}$; and a generalization of the H-theorem $dH/dt \leq 0$ is obtained.

The last three sections are devoted to polyatomic molecules, and aim to obtain generalized results for this case too. The key ingredient for doing so is, of course, an appropriately generalized *Stoßzahlansatz*. The formulation of this assumption is essentially the same as the one given in his paper on poly-atomic molecules 1871a, which was later shown wrong and corrected by Lorentz. I will not go into this issue (cf. [Lorentz, 1887; Boltzmann, 1887b; Tolman, 1938]).

An interesting passage occurs at the very end of the paper, where he expands on the relationship between H and entropy. He considers a monatomic gas in equilibrium. The stationary distribution of state is given as:

$$f^*(\vec{r}, \vec{v}) = V^{-1} \left(\frac{3m}{4\pi T} \right)^{3/2} \exp(\frac{-3mv^2}{4T}) \tag{53}$$

where V is the volume of the container. (Note that in comparison with (27), Boltzmann adopts units for temperature that make $k = 2/3$.) He shows that

$$H[f^*] := \int f^* \log f^* dx dv = -N \log V \left(\frac{4\pi T}{3m} \right)^{3/2} - \frac{3}{2} N; \tag{54}$$

which agrees (assuming $S = -kNH[f^*]$) with the thermodynamical expression for the ideal gas (16) up to an additive constant. A similar result holds for the polyatomic gas.

Remarks and problems

1. The role of probability. As we have seen, the H-theorem formed the basis of a renewed claim by Boltzmann to have obtained a theorem corresponding to the full Second Law (i.e. including both parts) at least for gases. A main difference from his 1866 claim, is that he now strongly emphasizes the role of probability calculus in his derivation. It is clear that the conception of probability expounded here is thoroughly frequentist and that he takes 'the laws of probability' as empirical statements. Furthermore, probabilities can be fully expressed in mechanical terms: the probability distribution f is nothing but the relative number of particles whose molecular states lie within certain limits. Thus, there is no conflict between his claims that on the one hand, "the problems of the mechanical theory of heat are really problems in probability calculus" and that the probabilities themselves are derived on the basis of the equations of motion alone, on the other hand. Indeed, it seems to me that Boltzmann's emphasis on the crucial role of probability in this paper is only intended to convey that probability theory provides a particularly useful and appropriate language for discussing mechanical problems in gas theory. There is no indication in this paper yet that probability theory could play a role by furnishing assumptions of a non-mechanical nature, i.e., independent of the equations of motion (cf. [Boltzmann and Nabl, 1904, p. 520]).

2. The role of the Stoßzahlansatz. Note that Boltzmann stresses the generality, rigour and "analyticity" of his proof. He puts no emphasis on the special assumptions that go into the argument. Indeed, the *Stoßzahlansatz*, later identified as the key assumption that is responsible for the time-asymmetry of the H-theorem, is announced as follows

> The determination [of the number of collisions] can only be obtained in a truly tedious manner, by consideration of the relative velocities of both particles. But since this consideration has, apart from its tediousness, not the slightest difficulty, nor any special interest, and because the result is so simple that one might almost say it is self-evident I will only state this result." [Boltzmann, 1909, I, p. 32].)

It thus seems natural that Boltzmann's contemporaries must have understood him as claiming that the H-theorem followed necessarily from the dynamics of the mechanical gas model.[31] I can find no evidence in the paper that he intended this claim to be read with a pinch of salt, as [von Plato, 1991, p.. 81] has argued.

[31] Indeed this is *exactly* how Boltzmann's claims were understood. For example, the recommendation written in 1888 for his membership of the Prussian Academy of Sciences mentions as his main feat that Boltzmann had proven that, whatever its initial state, a gas must necessarily approach the Maxwellian distribution [Kirsten and Körber, 1975, p.109].

Is there then no evidence at all for von Plato's reading of the paper? Von Plato refers to a passage from Section II, where Boltzmann repeats the previous analysis by assuming that energy can take on only discrete values, and replacing all integrals by sums. He recovers, of course, the same conclusion, but now adds a side remark, which touches upon the case of non-uniform gases:

> Whatever may have been the initial distribution of states, there is one and only one distribution which will be approached in the course of time. [...] This statement has been proved for the case where the distribution of states was already initially uniform. It must also be valid when this is not the case, i.e. when the molecules are initially distributed in such a way that in the course of time they mix among themselves more and more, so that after a very long time the distribution of states becomes uniform. This will always be the case, with the exception of very special cases, e.g. when all molecules were initially situated along a straight line, and were reflected by the walls onto this line [Boltzmann, 1909, I, p. 358].

It is this last remark that, apparently, led to the view that after all Boltzmann did already conceive of exceptions to his claims. However, I should say that this passage does not convince me. True enough, Boltzmann in the above quote indicates that there are exceptions. But he mentions them only in connection with an *extension* of his results to the case when the gas is not initially uniform, i.e. when condition (2) above is dropped. There can be no doubt that under the assumption of the conditions (1) and (2), Boltzmann claimed the rigorous validity of the H-theorem. (Curiously, his more systematic treatment of the non-uniform gas (Section III of 1872) does not mention any exception to the claim that "H can only decrease" [Boltzmann, 1909, I p. 362].

As a matter of fact, when Loschmidt formulated the objection, it happened to be by means of an example of a non-uniform gas (although nothing essential depended on this). Thus, if Boltzmann had in 1872 a "full hand against his future critics", as von Plato claims, one would expect his reply to Loschmidt's objection to point out that Loschmidt was correct but that he had already anticipated the objection. Instead, he accused Loschmidt of a fallacy (see paragraph 4.3 below).

But apart from the historical issue of whether Boltzmann did or did not envisage exceptions to his H-theorem, it seems more important to ask what kind of justification Boltzmann might have adduced for the *Stoßzahlansatz*. An attempt to answer this question must be somewhat speculative, since, as we have seen, Boltzmann presented the assumption as "almost self-evident" and "having no special interest", and hence presumably as not in need of further explanation. Still the following remarks may be made with some confidence.

First, we have seen that Maxwell's earlier usage of the assumption was never far away from an argument from insufficient reason. Thus, in his approach, one could think of the *Stoßzahlansatz* as expressing that we have no reason to expect any influence or correlation between any pair of particles that are about to collide.

The assumption would then appear as a probabilistic assumption, reflecting a 'reasonable judgment', independent from mechanics.

In contrast, Boltzmann's critique of Maxwell's approach (cf. footnote 16) suggests that he did not buy this argument for insufficient reason. But since the *Stoßzahlansatz* clearly cannot be conceived of as an assumption about dynamics — like the ergodic hypothesis — , this leaves only the option that it must be due to a special assumption about the mechanical state of the gas. Indeed, in the years 1895-6, when Boltzmann acknowledged the need for the *ansatz* in the proof of his H-theorem more explicitly — referring to it as "Assumption A" [Boltzmann, 1895] or "the hypothesis of molecular disorder" [Boltzmann, 1896] — , he formulated it as an assumption *about* the state of the gas.

Yet, even in those years, he would also formulate the hypothesis as expressing that "haphazard governs freely" [Boltzmann, 1895, Abh. III, p. 546] or "that the laws of probability are applicable for finding the number of collisions" [Boltzmann, 1895b]. Similarly, he describes states for which the hypothesis fails as contrived "so as to intentionally violate the laws of probability"[Boltzmann, 1896, §3]. However, I think these quotations should not be read as claims that the *Stoßzahlansatz* was a consequence of probability theory itself. Rather, given Boltzmann's empirical understanding of "the laws of probability", they suggest that Boltzmann thought that, as a matter of empirical fact, the assumption would 'almost always' hold, even if the gas was initially very far from equilibrium.

3. The H-theorem and the Second Law. Note that Boltzmann misconstrues, or perhaps understates, the significance of his results. Both the Boltzmann equation and the H-theorem refer to a body of gas in a fixed container that evolves in isolation from its environment. There is no question of heat being exchanged by the gas during a process, let alone in an irreversible cyclic process. His comparison in the quotation on page 965 with Clausius' integral $\int dQ/T$ (i.e. $\oint đQ/T$ in equation (18) above) is therefore really completely out of place.

The true import of Boltzmann's results is rather that they provide (i) a generalization of the entropy concept to non-equilibrium states,[32] and (ii)a claim that this non-equilibrium entropy $-kH$ increases monotonically as the isolated gas evolves for non-equilibrium towards an equilibrium state. The relationship with the Second Law is, therefore, somewhat indirect: On the one hand, Boltzmann proves much more than was required, since the second law does not speak of non-equilibrium entropy, nor of monotonic increase; on the other hand it proves also less, since Boltzmann does not consider the increase of entropy in general adiabatic processes.

[32]Boltzmann emphasized that his expression for entropy should be seen as an *extension* of thermodynamic entropy to non-equilibrium states in [1877b, (Boltzmann, 1909, II, p. 218); 1896, §5] . Of course there is no guarantee that this generalization is the *unique* candidate for a non-equilibrium entropy.

4.3 Boltzmann (1877a): the reversibility objection

According to [Klein, 1973], Boltzmann seemed to have been satisfied with his treatments of 1871 and 1872 and turned his attention to other matters for a couple of years. He did come back to gas theory in 1875 to discuss an extension of the Boltzmann equation to gases subjected to external forces. But this paper does not present any fundamental changes of thought. (However, it does contain some further elucidation, for example, it mentions for the first time that the derivation of the Boltzmann equation requires that the gas is so dilute that collisions between three or more particles simultaneously can be ignored).

However, the 1875 paper did contain a result which, two years later, led to a debate with Loschmidt. Boltzmann showed that (52) implied that a gas in equilibrium in an external force field (such as the earth's gravity) should have the same average kinetic energy at all heights and therefore, a uniform temperature; while its pressure and density would of course vary with height. This conclusion conflicted with the intuition that when molecules travel upwards, they must do work against the gravitational field, and pay for this by having a lower kinetic energy at greater heights.

Now Boltzmann (1875) was not the first to reach the contrary result, and Loschmidt was not the first to challenge it. Maxwell and Guthrie entered into a debate on the very same topic in 1873. But actually their main point of contention need not concern us very much. The discussion between Loschmidt and Boltzmann is particularly important for quite another issue, which Loschmidt only introduced as an side remark. Considering a gas container in a homogeneous gravitational field, Loschmidt discussed a situation where initially all atoms except one lie at rest at the bottom of the container. The single moving atom could then, by collisions, stir the others and send them into motion until a "stationary state", characterized by the Maxwell distribution, is obtained. He continues

> By the way, one should be careful about the claim that in a system in which the so-called stationary state has been achieved, starting from an arbitrary initial state, this average state can remain intact for all times. I believe, rather, that one can make this prediction only for a short while with full confidence.
>
> Indeed, if in the above case, after a time τ which is long enough to obtain the stationary state, one suddenly assumes that the velocities of all atoms are reversed, we would obtain an initial state that would appear to have the same character as the stationary state. For a fairly long time this would be appropriate, but gradually the stationary state would deteriorate, and after passage of the time τ we would inevitable return to our initial state: only one atom has absorbed all kinetic energy of the system [...], while all other molecules lie still on the bottom of the container.
>
> Obviously, in every arbitrary system the course of events must be become retrograde when the velocities of all its elements are reversed

[Loschmidt, 1876, p. 139].

Boltzmann's response (1877a)

Boltzmann's response to Loschmidt is somewhat confusing. On the one hand, he acknowledges that Loschmidt's objection is "quite ingenious and of great significance for the correct understanding of the Second Law." However, he also brands the objection as a "fallacy" and a "sophism".[33] But then, two pages later again, the argument is "of the greatest importance since it shows how intimately connected are the Second Law and probability theory."

The gist of the response is this. First, Boltzmann captures the essential core of the problem in an admirably clear fashion:

> "Every attempt to prove, from the nature of bodies and the laws of interaction for the forces they exert among each other, without any assumption about initial conditions, that
>
> $$\int \frac{dQ}{T} \leq 0 \tag{55}$$
>
> must be in vain" [Boltzmann, 1909, II. p.119–121].

The point raised here is usually known as the *reversibility objection.* And since the H-theorem (which only received this name in the 1890s) was presented in 1872 as a general proof that $\int \frac{dQ}{T} \leq 0$ (cf. the long quotation on page 965), it would imply that this theorem was invalid. Boltzmann aims to show, however, that this objection is a fallacy. His argument might be dissected into 5 central points.

1. Conceding that the proof cannot be given. Boltzmann says that a proof that every distribution must with absolute necessity evolve towards a uniform distribution cannot be given, claiming that this fact "is already taught by probability theory". Indeed, he argues, even a very non-uniform distribution of state is, although improbable to the highest degree, not impossible. Thus, he admits that there are initial states for which H increases, just as well as those for which H decreases. This admission, of course, is hard to rhyme with his professed purpose of showing that it is fallacious to conclude that some assumption about the initial state would be needed.

Note that this passage announces a major conceptual shift. Whereas the 1872 paper treated the distribution of state f_t as if it *defines* probability (i.e. of molecular velocities), this time the distribution of states is itself something which can be to a higher or lesser degree "probable". That is: probabilities are *attributed*

[33]The very fact that Boltzmann called this conclusion — which by all means and standards is *correct* — a fallacy shows, in my opinion, that he had not anticipated the objection. In fact, how much Boltzmann had yet to learn from Loschmidt's objection is evident when we compare this judgment to a quotation from his *Lectures on Gas Theory* [1898, p. 442]: "this one-sidedness [of the H-theorem] lies uniquely and solely in the initial conditions."

to distributions of state, i.e. the distribution of state itself is treated as a random variable. This shift in viewpoint became more explicit in his (1877b); as we will discuss in section 4.4 below.

2. Rethinking the meaning of "probability". Boltzmann argues that every distribution of state, whether uniform or non-uniform, is equally improbable. But there are "infinitely many" more uniform distributions of state than non-uniform distributions. Here we witness another conceptual shift. In (1872), the term "distribution of state" referred to the function $f(\vec{v})$ or $f(\vec{r}, \vec{v})$, representing the relative numbers of molecules with various molecular states. In that sense, there would, of course, only be a *single* uniform distribution of state: the Maxwellian distribution function (53). But since Boltzmann now claims there are many, he apparently uses the term "distribution of state" to denote a much more detailed description, that includes the velocity and position of every individual molecule, so that permutations of the molecules yield a different distribution of state. That is, he uses the term in the sense of what we would nowadays call a microstate, and what he himself would call a "*Komplexion*" a few months later in his (1877b) — on which occasion he would reserve the name 'distribution of state' for the macrostate.

Note that Boltzmann assumes every Komplexion to be equally probable (or improbable) so that the probability of a particular distribution of state is determined by the relative numbers. Indeed he remarks that it might be interesting to calculate the probabilities of state distributions by determining the ratio of their numbers; this suggestion is also worked out in his subsequent paper of 1877b.

This, indeed, marks another conceptual change. Not only are probabilities attributed to distributions of state instead of being defined by them; they are determined by an equiprobability assumption. Boltzmann does not explicitly motivate the assumption. In view of the discussion in paragraph 3.1, one might conjecture that he must have had something like Laplace's principle of insufficient reason in mind, which makes any two cases which, according to our information are equally possible, also equally probable. But this would indicate an even larger conceptual change; and not just because Boltzmann is broadly a frequentist concerning probability. Also, the principle of insufficient reason, or any similar assumption, makes sense only from the view point that probability is a non-mechanical notion: it reflects our belief or information about a system. I cannot find any evidence that he accepted this idea. Of course it is also possible to conjecture that he silently fell back upon the ergodic hypothesis. But this conjecture also seems unlikely, given his avoidance of the hypothesis since 1871.

3. A claim about evolutions. Boltzmann says: "Only from the fact that there are so many more uniform than non-uniform distributions of state [i.e.: microstates] follows the larger probability that the distribution will become uniform in the course of time" (p. 120). More explicitly, he continues:

> [...] one can prove that infinitely many more initial states evolve after a long time towards a more uniform distribution of states than to a less uniform one, and that even in that latter case, these states will become uniform after an even longer time [Boltzmann, 1909, II, p. 120][34]

Note that this is a claim about evolutions of microstates. In fact, it is the first case of what the Ehrenfests later called a *statistical H-theorem*, but what is perhaps better called a *statistical reading* of the H-theorem, since in spite of Boltzmann's assertion, no proof is offered.

4. The (im)probability of Loschmidt's initial state. Boltzmann maintains that the initial conditions considered by Loschmidt only have a minute probability. This is because it is obtained by a time evolution and velocity reversal of a non-uniform microstate. Since both time evolution and velocity reversal are one-to-one mappings (or more to the point: they preserve the Liouville measure), these operations should not affect the number or probability of states. Hence, the probability of Loschmidt's state is equal to that of the special non-uniform state from which it is constructed. But by point 2 above, there are infinitely many more uniform states than non-uniform states, so the probability of Loschmidt's state is extraordinarily small.

5. From (im)probability to (im)possibility. The final ingredient of Boltzmann's response is the claim that whatever has an extraordinarily small probability is practically impossible.

The conclusion of Boltzmann's argument, based on these five points, is that the state selected by Loschmidt may be considered as practically impossible. Note that this is a completely static argument; i.e., its logic relies merely on the points 1,2,4 and 5, and makes no assumption about evolutions, apart from the general feature that the dynamical evolution conserves states (or measure). Indeed, point 3, i.e. the statistical reading of the H-theorem, is not used in the argument.

As a consequence, the argument, although perfectly consistent, shows more than Boltzmann can possibly have wanted. The same reasoning that implies Loschmidt's initial state can be ignored, also excludes other non-uniform states. In particular, the same probability should be assigned to Loschmidt's initial state *without* the reversal of velocities. But that state *can* be produced in the laboratory, and, presumably, should not be considered as practically impossible. Indeed, if we adopt the rule that all non-uniform states are to be ignored on account of their low probability, we end up with a consideration of uniform states only, i.e. the theory would be reduced to a description of equilibrium, and the H-theorem reduced to $dH/dt = 0$, and any time-asymmetry is lost.

This, surely, is too cheap a victory over Loschmidt's objection. What one would like to see in Boltzmann's argument is a greater role for assumptions about the time evolution in order to substantiate his statistical reading of the H-theorem.

[34]The clause about 'the latter case' is absent in the translation by [Brush, 2003, p. 366].

Summing up: From this point on, we shall see that Boltzmann emphasizes even more strongly the close relations between the Second Law and probability theory. Even so, it is not always clear what these relations are exactly. Further, one may question whether his considerations of the probability of the initial state hit the nail on the head. Probability theory is equally neutral to the direction of time as is mechanics.

The true source of the reversibility problem was only identified by [Burbury, 1894a] and [Bryan, 1894] after Boltzmann's lecture in Oxford, which created a intense debate in the columns of *Nature*. They pointed out that the *Stoßzahlansatz* already contained a time-asymmetric assumption.

Indeed, this assumption requires that the number of collisions of the kind $(\vec{v}_1, \vec{v}_2) \longrightarrow (\vec{v}_1{}', \vec{v}_2{}')$ is proportional to the product $f(\vec{v}_1)f(\vec{v}_2)$ where, $\vec{v}_1, \vec{v}_2$ are the velocities *before* the collisions. If we would replace this by the requirement that the number of collisions is proportional to the product for the velocities $\vec{v}_1{}', \vec{v}_2{}'$ *after* the collision, we would obtain, by a similar reasoning, $dH/dt \geq 0$. The question is now, of course, why we should prefer one assumption above the other, without falling into some kind of double standard. (I refer to [Price, 1996] for a detailed discussion of this danger.) One thing is certain, and that is that any such preference cannot be obtained from mechanics and probability theory alone.

4.4 Boltzmann (1877b): the combinatorial argument

Boltzmann's next paper (1877b) is often seen as a major departure from the conceptual basis employed in his previous work. Indeed, the conceptual shifts already indicated implicitly in his reply to Loschmidt become in this article explicit. Indeed, according to [ter Haar, 1955, p. 296] and [Klein, 1973, p. 83], it is this paper that marks the transition from kinetic theory to statistical mechanics. Further, the paper presents the famous link between entropy and 'probability' that later became known as "Boltzmann's principle", and was engraved on his tombstone as "$S = k \log W$".

Boltzmann's begins the paper by stating that his goal is to elucidate the relationship between the Second Law and probability calculus. He notes he has repeatedly emphasized that the Second Law is related to probability calculus. In particular he points out that the 1872 paper confirmed this relationship by showing that a certain quantity [i.e. H] can only decrease, and must therefore obtain its minimum value in the state of thermal equilibrium. Yet, this connection of the Second Law with probability theory became even more apparent in his previous paper (1877a). Boltzmann states that he will now solve the problem mentioned in that paper, of calculating the probabilities of various distributions of state by determining the ratio of their numbers.

He also announces that, when a system starts in an improbable state, it will always evolve towards more probable states, until it reaches the most probable state, i.e. that of thermal equilibrium. When this is applied to the Second Law, he says, "we can identify that quantity which is usually called entropy, with the probability of the state in question." And: "According to the present interpretation, [the Second Law] states nothing else but that the probability of the total state of a composite system always increases" [Abh. II, pp. 165-6]. Exactly how all this is meant, he says, will become clear later in the article.

The combinatorial argument

Succinctly, and rephrased in the Ehrenfests' terminology, the argument is as follows. Apart from Γ, the mechanical phase space containing the possible states x for the total gas system, we consider the so-called μ-space, i.e. the state space of a single molecule. For monatomic gases, this space is just a six-dimensional Euclidean space with $(\vec{r}, \vec{v})$ as coordinates. With each mechanical state x we can associate a collection of N points in μ-space; one for each molecule.

Now, partition μ-space into m disjoint cells: $\mu = \omega_1 \cup \ldots \cup \omega_m$. These cells are taken to be rectangular in the coordinates and of equal size. Further, it is assumed that the energy of each molecule in cell ω_i in has a value ϵ_i, depending only on i. For each x, henceforth also called the *microstate* (Boltzmann's term was the *Komplexion*), we define the *macrostate* or 'distribution of state' as $Z := (n_1, \ldots, n_m)$, with n_i the number of particles whose molecular state is in cell ω_i. The relation between macro- and microstate is obviously non-unique since many different microstates, e.g. obtained by permuting the molecules, lead to the same macrostate. One may associate with every given macrostate Z_0 the corresponding set of microstates:

(56) $\Gamma_{Z_0} := \{x \in \Gamma \;:\; Z(x) = Z_0\}.$

The phase space volume $|\Gamma_{Z_0}|$ of this set is proportional to the number of permutations of the particles that do not change the macrostate Z_0. Indeed, when the six-dimensional volume of the cells ω_i is $\delta\omega$, i.e., the same for each cell, the phase space volume of the set Γ_Z is

(57) $|\Gamma_Z| = \dfrac{N!}{n_1! \cdots n_m!}(\delta\omega)^N.$

Moreover, assuming that $n_i \gg 1$ for all i and using the Stirling approximation for the factorials, one finds

(58) $\ln \Gamma_Z \approx N \ln N - \sum_i n_i \ln n_i + N \ln \delta\omega.$

This expression is in fact proportional to a discrete approximation of the H-function. Indeed, putting

(59) $n_i = N f(\vec{r}_i, \vec{v}_i)\delta\omega$

where $(\vec{r}_i, \vec{v}_i)$ are the coordinates of a representative point in ω_i, we find

$$\sum_i n_i \ln n_i \quad = \quad \sum_i N f(\vec{r}_i, \vec{v}_i) \ln \Big(N f(\vec{r}_i, \vec{v}_i)\delta\omega \Big)\, \delta\omega$$

$$\approx \quad N \int f(\vec{r},\vec{v}) \Big(\ln f(\vec{r},\vec{v}) + \ln N + \ln \delta\omega \Big) \, d^3\vec{r} d^3\vec{v}$$

$$= \quad NH + N \ln N + N \ln \delta\omega; \tag{60}$$

and therefore, in view of (58):

$$-NH \approx \ln |\Gamma_Z|. \tag{61}$$

And since Boltzmann had already identified $-kNH$ with the entropy of a macrostate, one can also take entropy as proportional to the logarithm of the volume of the corresponding region in phase space. Today, $\ln |\Gamma_Z|$ is often called the *Boltzmann entropy.*

Boltzmann next considers the question for which choice of Z does the region Γ_Z have maximal size, under the constraints of a given total number of particles N, and a total energy E:

$$N = \sum_{i=1}^{m} n_i, \qquad E = \sum_{i=1}^{m} n_i \epsilon_i. \tag{62}$$

This problem can easily be solved with the Lagrange multiplier technique. Under the Stirling approximation (58) one finds

$$n_i = \mu e^{\lambda \epsilon_i}, \tag{63}$$

which is a discrete version of the Maxwell distribution. (Here, μ an λ are determined in terms of N and E by the constraints (62).)

Boltzmann proposes to take the macrostate with the largest volume as representing equilibrium. More generally, he also refers to these volumes as the "probability" or "permutability" of the macrostate. He therefore now expresses the Second Law as a tendency for the system to evolve towards ever more probable macrostates, until, in equilibrium, it has reached the most probable state.

Remarks and problems

1. The role of dynamics. In the present argument, no dynamical assumption has been made. In particular, it is not relevant to the argument whether the ergodic hypothesis holds, or how the particles collide. At first sight, it might seem that this makes the present argument more general than the previous one. Indeed, Boltzmann suggests at the end of the paper [Boltzmann, 1909, II p. 223] that the same argument might be applicable also to dense gases and even to solids.

However, it should be noticed that the assumption that the total energy can be expressed in the form $E = \sum_i n_i \epsilon_i$ where the energy of each particle depends only on the cell in which it is located, and not on the state of other particles is very strong. This can only be maintained, independently of the number N, if there is no interaction at all between the particles. The validity of the argument is thus really restricted to ideal gases (cf. [Uhlenbeck and Ford, 1963]).

2. The choice of cells. One might perhaps hope, at first sight, that the procedure of partitioning μ-space into cells is only a technical or didactic device and can be eliminated by finally taking a limit in which $\delta\omega \longrightarrow 0$; similar to the procedure of his 1872 paper. This hope is dashed because the expression (58) diverges. Indeed, the whole prospect of using combinatorics would disappear if we did not adopt a finite partition. But also the special choice to give all cells equal volume in position and velocity variables is not quite self-evident, as Boltzmann himself shows. In fact, before he develops the argument given here, his paper presents a discussion in which the particles are characterized by their energy instead of position and velocity. This leads him to carve up μ-space into cells of equal size $\delta\epsilon$ in energy. He then shows that the combinatorial argument *fails* to reproduce the desired Maxwell distribution for particles moving in 3 spatial dimensions.[35] This failure is then remedied [Boltzmann, 1909, II, p. 190] by switching to a choice of equally sized cells in $\delta\omega$ in position and velocity. The latter choice is apparently 'right', in the sense that leads to the desired result. However, since the choice clearly cannot be relegated to a matter of convention, it leaves open the question of justification.

Modern commentators are utterly divided in the search for a direction in which a motivation for the choice of the size of these cells can be found. Some argue that the choice should be made in accordance with the actual finite resolution of measuring instruments or human observation capabilities. The question whether these do in fact favour a partition into cells of equal phase space volume has hardly been touched upon. Others [Popper, 1982; Redhead, 1995] reject an appeal to observation capacities on the grounds that these would introduce a 'subjective' or 'anthropocentric' element into the explanation of irreversibility (see also [Jaynes, 1965; Grünbaum, 1973; Denbigh and Denbigh, 1985; Ridderbos, 2002]).

3. Micro versus macro. The essential step in the argument is the distinction between micro- and macrostates. This is indeed the decisive new element, that allowed Boltzmann a complete reinterpretation of the notion and role of probability.

In 1872 and before, the distribution of state f was *identified* with a probability (namely of a molecular state, cf. Remark 1 of paragraph 4.2). On the other hand, in the present work it, or its discrete analogue Z, is a description of the macrostate of the gas, to which a probability is *assigned*. Essentially, the role of the distribution of state has been shifted from defining a probability measure to being a stochastic variable. Its previous role is taken over by a new idea: Probabilities are not assigned to the particles, but to the macrostate of the gas as a whole, and measured by the corresponding volume in phase space.

Another novelty is that Boltzmann has changed his concept of equilibrium. Whereas previously the defining characteristic of equilibrium was its stationarity,

[35] The problem is that for an ideal gas, where all energy is kinetic, $\delta\epsilon \propto v\delta v$. On the other hand, for three-dimensional particles, $\delta\omega \propto v^2\delta v$. The function f derived from (59) and (63) thus has a different dependence on v in the two cases. As Boltzmann notes, the two choices are compatible for particles in two dimensions (i.e. discs moving in a plane).

in Boltzmann's new view it is conceived as the macrostate (i.e. a region in phase space) that takes up the largest volume. As a result, a system in a state of equilibrium need not remain there: in the course of time, the microstate of the system may fluctuate in and out of this equilibrium region. Boltzmann briefly investigated the probability of such fluctuations in his [Boltzmann, 1878]. Almost thirty years later, the experimental predictions for fluctuation phenomena by Einstein and Smoluchowski provided striking empirical successes for statistical mechanics.

4. But what about evolutions? Perhaps the most important issue is this. What exactly is the relation of the 1877b paper to Loschmidt's objection and Boltzmann's primary reply to it (1877a)? The primary reply (cf. paragraph 4.3) can be read as an announcement of two subjects of further investigation:

> From the relative numbers of the various distributions of state, one might even be able to calculate their probabilities. This could lead to an interesting method of determining thermal equilibrium [Boltzmann, 1909, II, p. 121]

This is a problem about equilibrium. The second announcement was that Boltzmann said "The case is completely analogous for the Second Law" [Boltzmann, 1909, II, p. 121]. Because there are so very many more uniform than non-uniform distributions, it should be extraordinarily improbable that a system should evolve from a uniform distribution of states to a non-uniform distribution of states. This is a problem about evolution (cf. point 3 of section 4.3). In other words, one would like to see that something like the statistical H-theorem actually holds.

Boltzmann's [1877b] is widely read as a follow-up to these announcements. Indeed, Boltzmann repeats the first quote above in the introduction of the paper [Boltzmann, 1909, II, p. 165], indicating that he will address this problem. And so he does, extensively. Yet he also states:

> Our main goal is not to linger on a discussion of thermal equilibrium, but to investigate the relations of probability with the Second Law of thermodynamics [Boltzmann, 1909, II, p. 166].

Thus, the main goal of [1877b] is apparently to address the problem concerning evolutions and to show how they relate to the Second Law. Indeed, this is what one would naturally expect since the reversibility objection is, after all, a problem concerned with evolutions. Even so, a remarkable fact is that the 1877b paper hardly ever touches its self-professed "main goal" at all. As a matter of fact, I can find only one passage in the remainder of the paper where a connection with the Second Law is mentioned.

It occurs in Section V [Boltzmann, 1909, II, p. 216-7]. After showing that in equilibrium states for monatomic gases the 'permutability measure' $\ln|\Gamma_Z|$ (for which Boltzmann's notation is Ω) is proportional to the thermodynamical entropy, up to an arbitrary additive constant, he concludes that, by choosing the constant appropriately:[36]

$$(64)\quad \int \frac{dQ}{T} = \frac{2}{3}\Omega \left[= \frac{2}{3} \ln |\Gamma_Z|\right]$$

and adds:

> It is known that when a system of bodies goes through reversible changes, the total sum of the entropies of all these bodies remains constant; but when there are among these processes also irreversible (nicht umkehrbar) changes, then the total entropy must necessarily increase. This follows from the familiar circumstance that $\int dQ/T$ is negative for an irreversible cyclic process. In view of (64), the sum of all permutability measures of all bodies $\sum \Omega$, or their total permutability measure, must also increase. Hence, permutability is a quantity which is, up to a multiplicative and additive constant, identical to entropy, but which retains a meaning also during the passage of an irreversible body [sic– read: "process"], in the course of which it continually increases [Boltzmann, 1909, II p.217]

How does this settle the problem about evolutions, and does it provide a satisfactory refutation of the reversibility objection? In the literature, there are at least four views about what Boltzmann's response actually intended or accomplished.

4α. **Relying on the separation between micro- and macroscales:** A view that has been voiced recently, e.g. by [Goldstein, 2001], is that Boltzmann had, by his own argument, adequately and straightforwardly explained why entropy should tend to increase. In particular, this view argues, the fact of the overwhelmingly large phase space volume of the set Γ_{eq} of all equilibrium phase points, compared to the set of non-equilibrium points already provides a sufficient argument.

> For a non-equilibrium phase point x of energy E, the Hamiltonian dynamics governing the motion x_t arising from x would have to be ridiculously special to avoid reasonably quickly carrying x_t into Γ_{eq} and keeping it there for an extremely long time — unless, of course x itself were ridiculously special [Goldstein, 2001, p. 6].

In fact, this view may lay some claim to being historically faithful. As we have seen, [Boltzmann, 1877a] did claim that the large probability for an evolution towards equilibrium did follow from the large differences in number of states.

The main difficulty with this view is that, from a modern perspective, it is hard to maintain that it is adequate. States don't evolve into other states just because there are more of the latter, or because they make up a set of larger

[36] Actually, equation (64) is the closest he got to the famous formula on his tombstone, since $\Omega = \ln W$, and Boltzmann adopts a temperature scale that makes $k = 2/3$.

measure. The evolution of a system depends only on its initial state and its Hamiltonian. Questions about evolution can only be answered by means of an appeal to dynamics, not by the measure of sets alone. To take an extreme example, the trajectory covered by x_t, i.e. the set $\{x_t \ : \ t \in \mathbb{R}\}$ is a set of measure zero anyway; and hence very special. By contrast, its complement, i.e. the set of states *not* visited by a given trajectory is huge: it has measure one. Certainly, we cannot argue that the system cannot avoid wandering into the set of states that it does not visit. Another example is that of a system of non-interacting particles, e.g., the ideal gas. In this case, all the energies of the individual particles are conserved, and because of these conserved quantities, the phase point can only visit a very restricted region of phase space.[37]

The lesson is, of course, that in order to obtain any satisfactory argument why the system should tend to evolve from non-equilibrium states to the equilibrium state, we should make some assumptions about its dynamics. In any case, judgments like "reasonable" or "ridiculous" remain partly a matter of taste. The reversibility objection is a request for mathematical proof (which, as the saying goes, is something that even convinces an unreasonable person).

4β. **Relying on the ergodic hypothesis:** A second, and perhaps the most well-known, view to this problem is the one supplied by the Ehrenfests. In essence, they suggest that Boltzmann somehow relied on the ergodic hypothesis in his argument.

It is indeed evident that if the ergodic hypothesis holds, a state will spend time in the various regions of the energy hypersurface in phase space in proportion to their volume. That is to say, during the evolution of the system along its trajectory, regions with a small volume, corresponding to highly non-uniform distributions of state are visited only sporadically, and regions with larger volume, corresponding to more uniform distributions of state more often.

This should also make it plausible that if a system starts out from a very small region (an improbable state) it will display a tendency to evolve towards the overwhelmingly larger equilibrium state. Of course, this 'tendency' would have to be interpreted in a qualified sense: the same ergodic hypothesis would imply that the system cannot stay inside the equilibrium state forever and thus there would necessarily be fluctuations in and out of equilibrium. Indeed, one would have to state that the tendency to evolve from improbable to probable states is itself a probabilistic affair: as something that holds true for most of the initial states, or for most of the time, or as some or other form of average behaviour. In short, we would then hopefully obtain some statistical version of the H-theorem. What exactly the statistical H-theorem should say remains an open problem in the Ehrenfests' point of view. Indeed they distinguish between several interpretations (the so-called 'concentration curve' and the 'bundle of H-curves' [Ehrenfest and Ehrenfest-Afanassjewa, 1912, p. 31–35]).

[37]It is somewhat ironic to note, in view of remark 1 above, that this is the only case compatible with Boltzmann's argument. This gives rise to Khinchin's "methodological paradox" (cf. 1019).

Now, it is undeniable that the Ehrenfests' reading of Boltzmann's intentions has some clear advantages. In particular, even though nobody has yet succeeded in proving a statistical H-theorem on the basis of the ergodic hypothesis, or on the basis of the assumption of metric transitivity (cf. paragraph 6.1, one might hope that some statistical version of the H-theorem is true.

One problem here is that the assumptions Boltzmann used in his paper are restricted to non-interacting molecules, for which the ergodic hypothesis is demonstrably false. But even more importantly, it is clear that Boltzmann did not follow this line of argument in 1877b at all. Indeed, he nowhere mentions the ergodic hypothesis. In fact he later commented on the relation between the 1877b paper and the ergodic hypothesis of 1868, saying:

> On that occasion [i.e. in (1877b)] ... I did not wish to touch upon the question whether a system is capable of traversing all possible states compatible with the equation of energy [Boltzmann, 1881a, Abh. II p. 572].

4γ**. Relying on the H-theorem:** A third point of view, one to which this author adhered until recently, is that, in (1877b) Boltzmann simply relied on the validity of the H-theorem of 1872. After all, it was the 1872 paper that proposed to interpret $-NH$ as entropy (modulo multiplicative and additive constants), on the basis of the alleged theorem that it could never decrease. The 1877b paper presents a new proposal, to link the entropy of a macrostate with $\ln|\Gamma_Z|$. But this proposal is motivated, if not derived, by showing that $\ln|\Gamma_Z|$ is (approximately) equal to $-NH$, as in (61), whose interpretation as entropy was established in (1872). It thus seems plausible to conjecture that Boltzmann's thinking relied on the results of that paper, and that the claim that states will evolve from improbable to probable states, i.e. that $\ln|\Gamma_Z|$ shows a tendency to increase in time, likewise relied on the H-theorem he had proved there.[38] The drawback of this reading is that it makes Boltzmann's response to the reversibility objection quite untenable. Since the objection as formulated in his (1877a) calls the validity of the H-theorem into question, a response that *presupposes* the validity of this theorem is of no help at all.

4δ**. Bypassing the H-theorem:** [Janssen, 2002] has a different reading. He notes: "In Boltzmann's 1877 paper the statement that systems never evolve from more probable to less probable states is presented only as a new way of phrasing the Second Law, not as a consequence of the H-theorem" (p. 13). Indeed, any explicit reference to the H-theorem is absent in the 1877b paper. However, what we are to make of this is not quite certain. The earlier paper (1877a) did not mention the theorem either, but only discussed "any attempt to prove that $\int \frac{dQ}{T} \leq 0$". Still, this is commonly seen as an implicit reference to what is now known as the

[38]The conjecture is supported by the fact Boltzmann's later exposition in 1896 is presented along this line.

H-theorem, but which did not yet have a particular name at that time. Indeed, the H-theorem itself was characterized in 1872 only as a new proof that $\int \frac{dQ}{T} \leq 0$ (cf. the quotation on page 965). So, the fact that the H-theorem is not explicitly mentioned in (1877b) is not by itself a decisive argument that he did not intend to refer to it.

Even so, the fact that he presented the increase of entropy as something which was well-known and did not refer to the 1872 paper at all, does make Janssen's reading plausible. So, perhaps Boltzmann merely relied on the empirical validity of the Second Law as a ground for this statement, and not at all on any proposition from kinetic theory of gases.[39] This, of course, would undermine even more strongly the point of view that Boltzmann had a statistical version of the H-theorem, or indeed any theorem at all, about the probability of time evolution.

The reversibility objection was not about a relationship between the phenomenological Second Law and the H-theorem, but about the relationship between the H-theorem and the mechanical equations of motion. So even though Janssen's reading makes Boltzmann's views consistent, it does not make the 1877b paper provide a valid answer to Loschmidt's objection.

4ϵ. **The urn analogy — victory by definition?** At the risk of perhaps overworking the issue, I also want to suggest a fifth reading. Boltzmann's (1877b) contains an elaborate discussion of repeated drawings from an urn. In modern terms, he considers a Bernoulli process, i.e., a sequence of independent identically distributed repetitions of an experiment with a finite number of possible outcomes. To be concrete, consider an urn filled with m differently labeled lots, and a sequence of N drawings, in which the lot i is drawn n_i times ($\sum_{i=1}^m n_i = N$). He represents this sequence by a "distribution of state" $Z = (n_1, \ldots, n_m)$. In this discussion, the probability of these distributions of state is at first identified with the (normalized) number of permutations by which Z can be produced. In other words

$$(65)\quad \text{Prob}(Z) \propto \frac{N!}{n_1! \cdots n_m!}.$$

But halfway this discussion [Boltzmann, 1909, II, p. 171], he argues that one can redefine probabilities in an alternative fashion, namely, as the relative frequency of occurrence during *later* drawings of a sequence of N lots. Thus, even when, on a particular trial, an improbable state Z occurred, we can still argue that on a later drawings, a more probable state will occur. Boltzmann speaks about the changes in Z during the consecutive repetitions as an *evolution.* He then says:

[39] Further support for this reading can be gathered from later passages. For example, [Boltzmann, 1897b] writes "Experience shows that a system of interacting bodies is always found 'initially' in an improbable state and will soon reach the most probable state (that of equilibrium). [Boltzmann, 1909, III, p. 607]. Here too, Boltzmann presents the tendency to evolve from improbable to more probable states as a fact of experience rather than the consequence of any theorem.

> The most probable distribution of state must therefore be defined as that one to which most [states] will evolve to [Boltzmann, 1909, II, p. 172].

Although he does not make the point quite explicitly, the discussion of urn drawings is undoubtedly meant as an analogy for the evolution of the distribution of state in a gas. Hence, it is not implausible that, in the latter case too, Boltzmann might have thought that *by definition* the most probable distribution of state is the one that most states will evolve to. And this, in turn, would mean that he regarded the problem about evolutions not as something to be proved, and that might depend on the validity of specific dynamical assumptions like the ergodic hypothesis or the *Stoßzahlansatz*, but as something already settled from the outset. This would certainly explain why Boltzmann did not bother to address the issue further.

Even so, this reading too has serious objections. Apart from the fact that it is not a wise idea to redefine concepts in the middle of an argument, the analogy between the evolution of an isolated gas and a Bernoulli process is shaky. In the first case, the evolution is governed by deterministic laws of motion; in the latter one simply avoids any reference to underlying dynamics by the stipulation of the probabilistic independence of repeated drawings. However, see paragraph 6.2.

To sum up this discussion of Boltzmann's answer to the reversibility objection: it seems that on all above readings of his two 1877 papers, the lacuna between what Boltzmann had achieved and what he needed to do to answer Loschmidt satisfactorily — i.e. to address the issue of the evolution of distributions of state and to prove that non-uniform distributions tend, in some *statistical* sense, to uniform ones, or to prove any other reformulation of the H-theorem — remains striking.

4.5 The recurrence objection

Poincaré

In 1890, in his famous treatise on the three-body problem of celestial mechanics, Poincaré derived what is nowadays called the recurrence theorem . Roughly speaking, the theorem says that for every mechanical system with a bounded phase space, almost every initial state of the system will, after some finite time, return to a state arbitrarily closely to this initial state, and indeed repeat this infinitely often.

In modern terms, the theorem can be formulated as follows:

Recurrence Theorem: Consider a dynamical system[40] $\langle \Gamma, \mathcal{A}, \mu, T \rangle$

[40]See section 6.1 for a definition of dynamical systems. But in short: Γ is a phase space, $\mathcal{A}$ a family of measurable subsets of Γ and T is a one-parameter continuous group of time evolutions $T_t : \Gamma \times \mathbb{R} \longrightarrow \Gamma$.

with $\mu(\Gamma) < \infty$. Let $A \in \mathcal{A}$ be any measurable subset of Γ,and define, for a given time τ, the set

(66) $B = \{x \ : \ x \in A \ \& \ \forall t \geq \tau : T_t x \notin A\}$

Then

(67) $\mu(B) = 0$.

In particular, for a Hamiltonian system, if we choose Γ to be the energy hypersurface Γ_E, take A to be a 'tiny' region in Γ_E, say an open ball of diameter ϵ in canonical coordinates, the theorem says that the set of points in this region whose evolution is such that they will, after some time τ, never return to region A, has measure zero. In other words, almost every trajectory starting within A will after any finite time we choose, later return to A.

Poincaré had already expressed his objections against the tenability of a mechanical explanation of irreversible phenomena in thermodynamics earlier (e.g. [Poincaré, 1889]). But armed with his new theorem, he could make the point even stronger. In his 1893, he argued that the mechanical conception of heat is in contradiction with our experience of irreversible processes. According to the English kinetic theories, says Poincaré:

> [t]he world tends at first towards a state where it remains for a long time without apparent change; and this is consistent with experience; but it does not remain that way forever, it the theorem cited above is not violated; it merely stays there for an enormously long time, a time which is longer the more numerous are the molecules. This state will not be the final death of the universe but a sort of slumber, from which it will awake after millions and millions of centuries.
>
> According to this theory, to see heat pass from a cold body into a warm one, it will not be necessary to have the acute vision, the intelligence and the dexterity of Maxwell's demon; it will suffice to have a little patience [Brush, 2003, p.380].

He concludes that these consequences contradict experience and lead to a "definite condemnation of mechanism" [Brush, 2003, p.381].

Of course, Poincaré's "little patience", even for "millions and millions of centuries" is a rather optimistic understatement. Boltzmann later estimated the time needed for a recurrence in 1 cc of air to be $10^{10^{19}}$ seconds (see below): utterly beyond the bounds of experience. Poincaré's claim that the results of kinetic theory are contradicted by experience is thus too hasty.

Poincaré's article does not seem to have been noticed in the contemporary German-language physics community — perhaps because he criticized English theories only. However, Boltzmann was alerted to the problem when a slightly different argument was put forward by Zermelo in 1896. The foremost difference is that in Zermelo's argument experience does not play a role.

Zermelo's argument

Zermelo (1896a) points out that for a Hamiltonian mechanical system with a bounded phase space, Poincaré's theorem implies that, apart from a set of singular states, every state must recur almost exactly to its initial state, and indeed repeat this recurrence arbitrarily often. As a consequence, for any continuous function F on phase space, $F(x_t)$ cannot be monotonically increasing in time, (except when the initial state is singular); whenever there is a finite increase, there must also be a corresponding decrease when the initial state recurs. (see [Olsen, 1993] for a modern proof of this claim) Thus, it would be impossible to obtain 'irreversible' processes. Along the way, Zermelo points out a number of options to avoid the problem.

1. Either we assume that the gas system has no bounded phase space. This could be achieved by letting the particles reach infinite distances or infinite velocities. The first option is however excluded by the assumption that a gas is contained in a finite volume. The second option could be achieved when the gas consists point particles which attract each other at small distances, (e.g. an $F \propto r^{-2}$ inter-particle attractive force can accelerate them toward arbitrarily high velocities.) However, on physical grounds one ought to assume that there is always repulsion between particles at very small distances.

2. Another possibility is to assume that the particles act upon each other by velocity-dependent forces. This, however would lead either to a violation of the conservation of energy or the law of action and reaction, both of which are essential to atomic theory.

3. The H-theorem holds only for those special initial states which are the exception to the recurrence theorem, and we assume that only those states are realized in nature. This option would be unrefutable, says Zermelo. Indeed, the reversibility objection has already shown that not all initial states can correspond to the Second Law. However, here we would have to exclude the overwhelming majority of all imaginable initial states, since the exceptions to the Recurrence Theorem only make up a set of total extension (i.e. in modern language: measure) zero. Moreover, the smallest change in the state variables would transform a singular state into a recurring state, and thus suffice to destroy the assumption. Therefore, this assumption "would be quite unique in physics and I do not believe that anyone would be satisfied with it for very long."

This leaves only two major options:

4. The Carnot-Clausius principle must be altered.[41]

5. The kinetic theory must be formulated in an essentially different way, or even be given up altogether.

Zermelo does not express any preference between these last two options. He concludes that his aim has been to explain as clearly as possible what can be proved rigorously, and hopes that this will contribute to a renewed discussion and

[41] By this term, Zermelo obviously referred to the Second Law, presumably including the Zeroth Law.

final solution of the problem.

I would like to emphasize that, in my opinion, Zermelo's argument is entirely correct. If he can be faulted for anything, it is only that he had not noticed that in his very recent papers, Boltzmann had already been putting a different gloss on the H-theorem.

Boltzmann's response

[Boltzmann, 1896b] response opens by stating that he had repeatedly pointed out that the theorems of gas are statistical. In particular, he says, he had often emphasized as clearly as possible that the Maxwell distribution law is not a theorem from ordinary mechanics and cannot be proven from mechanical assumptions.[42] Similarly, from the molecular viewpoint, the Second Law appears merely as a probability statement. He continues with a sarcastic remark:

> Zermelo's paper shows that my writings have been misunderstood; nevertheless it pleases me for it appears to be the first indication that these works have been noticed in Germany.[43]

Boltzmann agrees that Poincaré's recurrence theorem is "obviously correct", but claims that Zermelo's application of the theorem to gas theory is not. His counter argument is very similar to his (1895) presentation in *Nature*, a paper that Zermelo had clearly missed.

In more detail, this argument runs as follows. Consider a gas in a vessel with perfectly smooth and elastic walls, in an arbitrary initial state and let it evolve in the course of time. At each time t we can calculate $H(t)$. Further, consider a graph of this function, which Boltzmann called: *the H-curve*. In his second reply to Zermelo [Boltzmann, 1897a], he actually produced a diagram. A rough an modernized version of such an H-curve is sketched in Fig. 3.

Barring all cases in which the motion is 'regular', e.g. when all the molecules move in one common plane, Boltzmann claims the following properties of the curve:

(i). For most of the time, $H(t)$ will be very close to its minimum value, say $H_{\min}$. Moreover, whenever the value of $H(t)$ is very close to $H_{\min}$, the distribution of molecular velocities deviates only very little from the Maxwell distribution.

[42]This is, as we have seen, a point Boltzmann had been making since 1877. However, one might note that just a few years earlier, [Boltzmann, 1892], after giving yet another derivation of the Maxwell distribution (this time generalized to a gas of hard bodies with an arbitrary number of degrees of freedom that contribute quadratic terms to the Hamiltonian), had concluded: "I believe therefore that its correctness [i.e. of the Maxwell distribution law] as a theorem of analytical mechanics can hardly be doubted" [Boltzmann, 1909, III p.432]. But as we have seen on other occasions, for Boltzmann, statements that some result depended essentially on probability theory, and the statement that it could be derived as a mechanical theorem, need not exclude each other.

[43]Eight years earlier, Boltzmann had been offered the prestigious chair in Berlin as successor of Kirchhoff, and membership of the Prussian Academy. The complaint that his works did not draw attention in Germany is thus hard to take seriously.

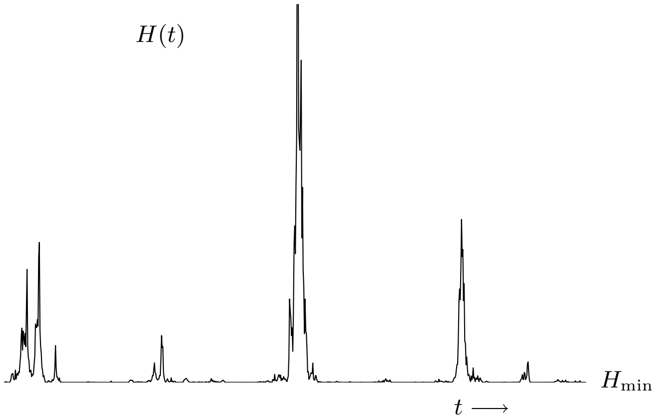


Figure 3. A (stylized) example of an H-curve

(ii). The curve will occasionally, but very rarely, rise to a peak or summit, that may be well above $H_{\min}$.

(iii). The probability of a peak decreases extremely rapidly with its height.

Now suppose that, at some initial time $t = 0$, the function takes a very high value H_0, well above the minimum value. Then, Boltzmann says, it will be enormously probable that the state will, in the course of time, approach the Maxwell distribution, i.e., $H(t)$ will decrease towards $H_{\min}$; and subsequently remain there for an enormously long time, so that the state will deviate only very little from the Maxwell distribution during vanishingly short durations. Nevertheless, if one waits even longer, one will encounter a new peak, and indeed, the original state will eventually recur. In a mathematical sense, therefore, these evolutions are periodic, in full conformity with Poincaré's recurrence theorem.

What, then, is the failure of Zermelo's argument? Zermelo had claimed that only very special states have the property of continually approaching the Maxwell distribution, and that these special states taken together make up an infinitely small number compared to the totality of possible states. This is incorrect, Boltzmann says. For the overwhelming majority of states, the H-curve has the qualitative character sketched above.

Boltzmann also took issue with (what he claimed to be Zermelo's) conclusion that the mechanical viewpoint must somehow be changed or given up. This conclusion would only be justified, he argues, if this viewpoint led to some consequence that contradicted experience. But, Boltzmann claims, the duration of the recurrence times is so large that no one will live to observe them.

To substantiate this claim about the length of the recurrence time, he presents, in an appendix an estimate of the recurrence time for 1 cc of air at room temperature and pressure. Assuming there are 10^9 molecules in this sample,[44] and choosing cells in the corresponding μ-space as six-dimensional cubes of width 10^{-9} m in (physical) space and 1 m/s in velocity space, Boltzmann calculates the number of different macrostates, i.e. the number of different ways in which the molecules can be distributed over these cells as (roughly) 10^{10^9}. He then assumes that, before a recurrence of a previous macrostate, the system has to pass through *all* other macrostates. Even if the molecules collide very often, so that the system changes its macrostate 10^{27} times per second, the total time it takes to go through this huge number of macrostates will still take $10^{10^9-27} \approx 10^{10^9}$ seconds. In fact, this time is so immensely large that its order of magnitude is not affected whether we express it in seconds, years, millennia, or what have you.

The upshot is, according to Boltzmann: if we adopt the view that heat is a form of motion of the molecules, obeying the general laws of mechanics, and assume that the initial state of a system is very unlikely, we arrive at a theorem which corresponds to the Second Law for all observed phenomena. He ends with another sarcasm:

> All the objections raised against the mechanical view of Nature are therefore empty and rest on errors. But whoever cannot overcome the difficulties, which a clear understanding of the theorems of gas theory poses, should indeed follow the advice of Mr Zermelo and decide to give up the theory completely. [Boltzmann, 1909, III p. 576].

Zermelo's reply

[Zermelo, 1896b] notes that Boltzmann's response confirms his views by admitting that the Poincaré theorem is correct and applicable to a closed system of gas molecules. Hence, in such a system, "all [sic] motions are *periodic* and not *irreversible* in the strict sense". Thus, kinetic gas theory cannot assert that there is a strict monotonic increase of entropy as the Second Law would require. He adds: "I think this general clarification was not at all superfluous" [Brush, 2003, p. 404].

Therefore, Zermelo argues, his main point had been conceded: there is indeed a conflict between thermodynamics and kinetic theory, and it remains a matter of taste which of the two is abandoned. Zermelo admits that observation of the Poincaré recurrences may well fall beyond the bounds of human experience. He points out (correctly) that Boltzmann's estimate of the recurrence time presupposes that the system visits *all* other cells in phase space before recurring to an initial state. This estimate is inconclusive, since the latter assumption is somewhat ad hoc. In general, these recurrence times need not "come out so 'comfortingly' large" [Brush, 2003, p. 405]. But, as I stressed before, the relation with experience simply was no issue in Zermelo's objection.

[44] Actually, modern estimates put the number of molecules in 1cc of air closer to 10^{19}, which would make Boltzmann's estimate for recurrence time even larger still, i.e. $10^{10^{19}}$.

The main body of Zermelo's reply is taken by an analysis of the justification of and consequences drawn from Boltzmann's assumption that the initial state is very improbable, i.e., that H_0 is very high. Zermelo argues that even in order to obtain an approximate or empirical analogue of the Second Law, as Boltzmann envisaged, i.e. an approach to a long-lasting, but not permanent equilibrium state, it would not suffice to show this result for one particular initial state. Rather, one would have to show that evolutions *always* take place in the same sense, at least during observable time spans.

As Zermelo understands it, Boltzmann does not merely assume that the initial state has a very high value for H, but also that, as a rule, the initial state lies on a maximum, or has just passed a maximum. If this assumption is granted, then it is obvious that one can only observe a decreasing flank of the H-curve. However, Zermelo protests, one could have chosen any time as the initial time. In order to obtain a satisfactorily general result, the additional assumption would thus have to apply at all times. But then the H-curve would have to consist entirely of maxima. But this leads to nonsense, Zermelo argues, since the curve cannot be constant. Zermelo concludes that Boltzmann's assumptions about the initial state are thus in need of further *physical* explanation.

Further, Zermelo points out that probability theory, by itself, is neutral with respect to the direction of time, so that no preference for evolutions in a particular sense can be derived from it. He also points out that Boltzmann apparently equates the duration of a state and its extension (i.e. the relative time spent in a region and the relative volume of that region in phase space). "I cannot find that he has actually *proved* this property" [Brush, 2003, p. 406].

Boltzmann's second reply

In his second reply 1897a, Boltzmann rebuts Zermelo's demand for a physical explanation of his assumptions about the initial state of the system with the claim that the question is not what will happen to an arbitrarily chosen initial state, but rather what will happen to a system in the present state of the universe.

He argues that one should depart from the (admittedly unprovable) assumption that the universe (or at least a very large part of the universe that surrounds us started in a very improbable state and still is in an improbable state. If one then considers a small system (e.g. a gas) that is suddenly isolated from the rest of the universe, there are the following possibilities: (i) The system may already be in equilibrium, i.e. H is close to its minimum value. This, Boltzmann says, is by far the most probable case. But among the few cases in which the system is not in equilibrium, the most probable case is (ii) that H will be on a maximum of the H-curve, so that it will decrease in both directions of time. Even more rare is the case in which (iii) the initial value of H will fall on a decreasing flank of the H curve. But such cases are just as frequent as those in which (iv) H falls on an increasing flank.[45]

[45]The Ehrenfests 1912 later added a final possible case (v): H may initially be on a local

Thus, Boltzmann's explanation for the claim that H is initially on a maximum is that this would be the most likely case for a system not in equilibrium, which isolated from the rest of the universe in its present state.

This occasion is perhaps the first time that Boltzmann advanced an explanation of his claims as being due to an assumption about initial state of the system, ultimately tied to an assumption about the initial conditions of the universe. Today, this is often called the *past-hypothesis* (cf. [Albert, 2000; Winsberg, 2004; Callender, 2004; Earman, 2006]).

He ends his reply with the observation that while the mechanical conception of gas theory agrees with the Clausius-Carnot conception [i.e. thermodynamics] in all observable phenomena, a virtue of the mechanical view is that it might eventually predict new phenomena, in particular for the motion of small bodies suspended in fluids. These prophetic words were substantiated eight years later in Einstein's work on Brownian motion.

However, he does not respond to Zermelo's requests for more definite proofs of the claims (1) –(3), or of the equality of phase space volume and time averages in particular. He bluntly states that he has thirty years of priority in measuring probabilities by means of phase space volume (which is true) and adds that he has always had done so (which is false). Even so, one cannot interpret this claim of Boltzmann as a rejection of the time average conception of probability. A few lines below, he claims that the most probable states will also occur most frequently, except for a vanishingly small number of initial states. He does not enter into a proof of this. Once again, this provides an instance where the Ehrenfests conjectured that Boltzmann might have had the ergodic hypothesis in the back of his mind.

Remarks

Boltzmann's replies to Zermelo have been recommended as "superbly clear and right on the money" [Lebowitz, 1999, p. S347]. However, as will clear from the above and the following remarks, I do not share this view. See also [Klein, 1973; Curd, 1982; Batterman, 1990; Cercignani, 1998; Brush, 1999; Earman, 2006] for other commentaries on the Zermelo-Boltzmann dispute.

1. The issues at stake It is clear that, in at least one main point of the dispute, Boltzmann and Zermelo had been talking past each other. When Zermelo argued that in the kinetic theory of gases there can be no continual approach towards a final stationary state, he obviously meant this in the sense of a limit $t \longrightarrow \infty$. But Boltzmann's reply indicates that he took the "approach" as something that is not certain but only probable, and as lasting for a very long, but finite time. His graph of the H-curve makes abundantly clear that $\lim_{t \longrightarrow \infty} \mathrm{H(t)}$ does not exist.

It is true that his statistical reading of the H-theorem, as laid down in the claims (1)–(3) above, was already explicit in (Boltzmann 1895), and thus Boltzmann could

minimum of the H-curve, so that it increases in both directions of time. But by a similar reasoning, that case is even less probable than the cases mentioned by Boltzmann.

claim with some justification that his work had been overlooked. But in fairness, one must note that, even in this period, Boltzmann was sending mixed messages to his readers. Indeed, the first volume of Boltzmann's *Lectures on Gas Theory*, published in 1896, stressed, much like his original [1872] paper on the H-theorem, the necessity and exceptionless generality of the H-theorem, adding only that the theorem depended on the assumption of molecular disorder (as he then called the *Stoßzahlansatz*):[46] "

> [T]he quantity designated as H can only decrease; at most it can remain constant.[...] The only assumption we have made here is that the distribution of velocities was initially 'molecularly disordered' and remains disordered. Under this condition we have therefore proved that the quantity called H can only decrease and that the distribution of velocities must necessarily approach the Maxwell distribution ever more closely [Boltzmann, 1896, § 5, p. 38].

Zermelo might not have been alone in presuming that Boltzmann had intended this last claim literally, and was at least equally justified in pointing out that Boltzmann's clarification "was not at all superfluous".

On the other hand, Boltzmann misrepresented Zermelo's argument as concluding that the mechanical view should be given up. As we have seen, Zermelo only argued for a *dilemma* between the strict validity of the kinetic theory and the strict validity of thermodynamics. Empirical matters were not relevant to Zermelo's analysis. Still, Boltzmann is obviously correct when he says that the objection does not yet unearth a conflict with experience. Thus, his response would have been more successful as a counter-argument to Poincaré than to Zermelo.

2. The statistical reading of the H-theorem. Another point concerns the set of claims (1)–(3) that Boltzmann lays down for the behaviour of the H-curve. Together, they form perhaps the most clearly stated and explicit form of the "statistical reading of the H-theorem" (cf remark 3 on page 972). Yet they only have a loose connection to the original theorem. It is unclear, for example, whether these claims still depend on the *Stoßzahlansatz*, the assumption that the gas is dilute, etc. It thus remains a reasonable question what argument we have for their validity. Boltzmann offers none. In his 1895 paper in *Nature*, he argued as if he had proved as much in his earlier papers, and added tersely: "I will not here repeat the proofs given in my papers" [Boltzmann, 1909, III p. 541]. But surely, Boltzmann never proved anything concerning the probability of the time evolution of H, and at this point there remains a gap in his theory. Of course, one might

[46]in his reply to Zermelo, Boltzmann claimed that his discussion of the H-theorem in the *Lectures on Gas theory* was intended under the explicitly emphasized assumption that the number of molecules was infinite, so that the recurrence theorem did not apply. However, I can find no mention of such an assumption in this context. On the contrary, the first occasion on which this latter assumption appears is in §6 on page 46 where it is introduced as "an assumption we shall make later", suggesting that the previous discussion did *not* depend on in it.

speculate on ways to bridge this gap; e.g. that Boltzmann implicitly and silently relied on the ergodic hypothesis, as the Ehrenfests suggested or in other ways, but I refrain from discussing this further. The most successful modern attempt so far to formulate and prove a statistical H-theorem has been provided by Lanford, see paragraph 6.4 below.

5 GIBBS' STATISTICAL MECHANICS

The birth of statistical mechanics in a strict sense, i.e. as a coherent and systematic theory, is marked by the appearance of J.W. Gibbs's book (1902) which carries this title: *Elementary Principles in Statistical Mechanics; developed with especial reference to the rational foundation of thermodynamics.* His point of departure is a general mechanical system governed by Hamiltonian equations of motion, whose (micro)states are represented by points in the mechanical phase space Γ.

Gibbs avoids specific hypotheses about the microscopic constitution of such a system. He refers to the well-known problem concerning the anomalous values of the specific heat for gases consisting of diatomic molecules (mentioned in footnote 10), and remarks:

> Difficulties of this kind have deterred the author from attempting to explain the mysteries of nature, and have forced him to be contented with the more modest aim of deducing some of the more obvious propositions relating to the statistical branch of mechanics [Gibbs, 1902, p. viii].

It is clear from this quote that Gibbs' main concern was with logical coherence, and less with the molecular constitution. (Indeed, only the very last chapter of the book is devoted to systems composed of molecules.) This sets his approach apart from Maxwell and Boltzmann.[47]

The only two ingredients in Gibbs' logical scheme are mechanics and probability. Probability is introduced here as an ingredient not reducible to the mechanical state of an individual system, but by means of the now familiar "ensemble":

> We may imagine a great number of systems of the same nature, but differing in the configurations and velocities which they have at a given instant, and differing not merely infinitesimally, but it may be so as to embrace every conceivable combination of configuration and velocities. And here we may set the problem, not to follow a particular system through its succession of configurations, but to determine how the whole number of systems will be distributed among the various conceivable configurations and velocities at any required time, when the distribution has been given for some one time [Gibbs, 1902, p. v].

[47]It also sets him apart from the approach of Einstein who, in a series of papers (1902, 1903, 1904) independently developed a formalism closely related to that of Gibbs, but used it as a probe to obtain empirical tests for the molecular/atomic hypothesis (cf. [Gearhart, 1990; Navarro, 1998; Uffink, 2006]).

and

> What we know about a body can generally be described most accurately and most simply by saying that it is one taken at random from a great number (ensemble) of bodies which are completely described. (p. 163)

Note that Gibbs is somewhat non-committal about any particular interpretation of probability. (Of course, most of the presently distinguished interpretations of probability were only elaborated since the 1920s, and we cannot suppose Gibbs to have pre-knowledge of those distinctions.) A modern frequentist (for whom a probability of an event is the frequency with which that event occurs in a long sequence of similar cases) will have no difficulty with Gibbs' reference to an ensemble, and will presumably identify that notion with von Mises' notion of a *Kollektiv*. On the other hand, authors like Jaynes who favour a subjectivist interpretation of probability (in which the probability of an event is understood as a state of knowledge or belief about that event) have emphasized that in Gibbs' approach the ensemble is merely 'imagined' and a tool for representing our knowledge.

The ensemble is usually presented in the form of a probability density function ρ over Γ, such that $\int_A \rho(x)dx$ is the relative number of systems in the ensemble whose microstate $x = (\vec{q}_1, \vec{p}_1; \ldots; \vec{q}_N, \vec{p}_N)$ lies in the region A. The evolution of an ensemble density ρ_0 at time $t = 0$ is dictated by the Hamiltonian equations of motion. In terms of the (formal) time evolution operator T_t, we get

(68) $\rho_t(x) = \rho_0(T_{-t}x)$

or, in differential form:

(69) $$\frac{\partial \rho_t(x)}{\partial t} = \{H, \rho\}$$

where $\{\cdot, \cdot\}$ denotes the Poisson bracket:

(70) $$\{H, \rho\} = \sum_{i=1}^{N} \frac{\partial H}{\partial \vec{q}_i}\frac{\partial \rho}{\partial \vec{p}_i} - \frac{\partial H}{\partial \vec{p}_i}\frac{\partial \rho}{\partial \vec{q}_i}$$

A case of special interest is that in which the density function is stationary, i.e.

(71) $$\forall t: \quad \frac{\partial \rho_t(x)}{\partial t} = 0.$$

This is what Gibbs calls the condition of *statistical equilibrium*. Gibbs notes that any density which can be written as a function of the Hamiltonian is stationary, and proceeds to distinguish special cases, of which the most important are:

(72) $$\rho_E(x) = \frac{1}{\omega(E)}\delta(H(x) - E) \quad \text{(microcanonical)}$$

(73) $$\rho_\theta(x) = \frac{1}{Z(\theta)}\exp(-H(x)/\theta) \quad \text{(canonical)}$$

(74) $$\rho_{\theta,\alpha}(x, N) = \frac{1}{N!Z(\theta,\alpha)}\exp(-H(x)/\theta + \alpha N) \quad \text{(grand-canonical)}$$

where $\omega(E)$, $Z(\theta)$ and $Z(\theta, \alpha)$ are normalization factors. In the following I will mainly discuss the canonical and microcanonical ensembles.

5.1 Thermodynamic analogies for statistical equilibrium

As indicated by the subtitle of the book, Gibbs' main goal was to provide a 'rational foundation' for thermodynamics. He approaches this issue quite cautiously, by pointing out certain analogies between relations holding for the canonical and microcanonical ensembles and results of thermodynamics. At no point does Gibbs claim to have reduced thermodynamics to statistical mechanics.

The very first analogy noticed by Gibbs is in the case of two systems, A and B put into thermal contact. This is modeled in statistical mechanics by taking the product phase space, $\Gamma_{AB} = \Gamma_A \times \Gamma_B$, and a Hamiltonian $H_{AB} = H_A + H_B + H_{\text{int}}$. If both A and B are described by canonical ensembles and if H_{int} is 'infinitely small' compared to the system Hamiltonian, then the combined system will be in statistical equilibrium if $\theta_A = \theta_B$. This, he says, "is entirely analogous to . . . the corresponding case in thermodynamics" where "the most simple test of the equality of temperature of two bodies is that they remain in thermal equilibrium when brought into thermal contact" (ibid. p. 37). Clearly, Gibbs invites us to think of statistical equilibrium as analogous to thermal equilibrium, and θ as the analogue of the temperature of the system.[48]

A second point of analogy is in reproducing the 'fundamental equation' (23) of thermodynamics:

$$(75)\quad dU = TdS + \sum_i F_i da_i$$

where a_i are the so-called external parameters (e.g. volume) and F_i the associated generalized forces (e.g. minus the pressure). For the canonical ensemble, Gibbs derives a relation formally similar to the above fundamental equation:[49]

$$(76)\quad d\langle H\rangle = \theta d\sigma - \sum_i \langle A_i\rangle da_i.$$

Here, $\langle H\rangle$ is the expectation value of the Hamiltonian in the canonical ensemble, θ the modulus of the ensemble, σ the so-called Gibbs entropy of the canonical distribution:

$$(77)\quad \sigma[\rho_\theta] = -\int \rho_\theta(x) \ln \rho_\theta(x) dx,$$

a_i are parameters in the form of the Hamiltonian and the $\langle A_i\rangle = \langle \frac{\partial H}{\partial a_i}\rangle$ represent

[48]A more elaborate discussion of the properties of the parameter θ and their analogies to temperature, is in Einstein (1902). That discussion also addresses the transitivity of thermal equilibrium, i.e. the Zeroth Law of thermodynamics (cf. paragraph 2).

[49]See [Uhlenbeck and Ford, 1963; van Lith, 2001b] for details.

the 'generalized forces'.[50] The equation suggests that the canonical ensemble averages might serve as analogues of the corresponding thermodynamic quantities, and θ and σ as analogues of respectively temperature and entropy.[51]

Note the peculiarly different role of θ and σ in (76): these are not expectations of phase space functions, but a parameter and a functional of the ensemble density ρ_θ. This has a significant conceptual implication. The former quantities may be thought of as averages, taken over the ensemble of some property possessed by each individual system in the ensemble. But for temperature θ and entropy σ, this is not so. In the case of θ one can diminish this contrast — at least when H is the sum of a kinetic and a potential energy term and the kinetic part is quadratic in the momenta, i.e. $H = \sum_i \alpha_i p_i^2 + U(q_1, \ldots q_n)$ — because of the well-known equipartition theorem. This theorem says that θ equals twice the expected kinetic energy for each degree of freedom:

$$(78)\quad \frac{\theta}{2} = \alpha_i \langle p_i^2 \rangle_\theta .$$

Thus, in this case, one can find phase functions whose canonical expectation values are equal to θ, and regard the value of such a function as corresponding to the temperature of an individual system.[52] But *no* function χ on phase space exists such that

$$(79)\quad \sigma[\rho_\theta] = \langle \chi \rangle_\theta \quad \text{for all } \theta.$$

Thus, the Gibbs entropy cannot be interpreted as an average of some property of the individual members of the ensemble.

The next question is whether a differential equation similar to (76) can be obtained also for the microcanonical ensemble. In this case, it is natural to consider the same expressions $\langle A_i \rangle$ and $\langle H \rangle$ as above, but now taken as expectations with respect to the microcanonical ensemble, so that obviously $\langle H \rangle_{\rm mc} = E$. The problem is then to find the microcanonical analogies to T and S. [Gibbs, 1902, p. 124–128, 169–171] proposes the following:

$$(80)\qquad T \longleftrightarrow \left(\frac{d \ln \Omega(E)}{dE}\right)^{-1},$$

$$(81)\qquad S \longleftrightarrow \ln \Omega(E),$$

where

$$(82)\qquad \Omega(E) := \int_{H(x) \leq E} dp_1 \ldots dq_n$$

[50] A more delicate argument is needed if one wishes to verify that $-\langle \frac{\partial H}{\partial V} \rangle$ can really be identified with pressure, i.e. the average force per unit area on the walls of the container. Such an argument is given by [Martin-Löf, 1979, p. 21–25]

[51] A crucial assumption in this derivation is that the differential expressions represent infinitesimal elements of quasistatic processes during which the probability density always retains its canonical shape. This assumption is in conflict with a dynamical evolution [van Lith, 2001b, p. 141].

[52] For proposals of more generally defined phase functions that can serve as an analogy of temperature, see [Rugh, 1997; Jepps *et al.*, 2000].

is known as the integrated structure function.

Remarkably, in a later passage, [Gibbs, 1902, p. 172–178] also provides a second pair of analogies to temperature and entropy, namely:

$$T \longleftrightarrow \left(\frac{d\ln\omega(E)}{dE}\right)^{-1} \tag{83}$$

$$S \longleftrightarrow \ln\omega(E), \tag{84}$$

where ω is the structure function

$$\omega(E) = \frac{d\Omega(E)}{dE} = \int_{H(x)=E} dx.$$

For this choice, the relation (75) is again reproduced. Thus, there appears to be a variety of choice for statistical mechanical quantities that may serve as thermodynamic analogue. Although Gibbs discussed various pro's and con's of the two sets, — depending on such issues as whether we regard the energy or the temperature as an independent variable, and whether we prefer expected values of most probable values — he does not reach a clear preference for one of them. (As he put it, system (80,81) is the more natural, while system (83,84) is the simpler of the two.) Still, Gibbs argued (ibid., p. 183) that the two sets of analogies will approximately coincide for a very large number degrees of freedom. Nevertheless, this means there remains an underdetermination in his approach that one can hope to avoid only in the thermodynamic limit.

The expressions (81) and (84) are also known as the 'volume entropy' and the 'surface entropy'. In modern textbooks the latter choice has been by far the most popular, perhaps because it coincides with the Gibbs entropy for the microcanonical ensemble: $\sigma[\rho_E] = \ln\omega(E)$. However, it has been pointed out that there are also general theoretical reasons to prefer the volume entropy (81), in particular because it is, unlike the surface entropy, an adiabatic invariant (see [Hertz, 1910; Rugh, 2001; Campisi, 2005]).

Of course, all of this is restricted to (statistical) equilibrium. In the case of non-equilibrium, one would obviously like to obtain further thermodynamical analogies that recover the approach to equilibrium (the 'Minus First Law', cf. p. 939) and an increase in entropy for adiabatic processes that start and end in equilibrium, or even to reproduce the kinetic equations on a full statistical mechanical basis. What Gibbs had to say on such issues will be the subject of the paragraphs 5.3 and 5.4.

But Gibbs also noted that a comparison of temperature and entropy with their analogies in statistical mechanics "would not be complete without a consideration of their differences with respect to units and zeros and the numbers used for their numerical specification" [Gibbs, 1902, p.183]. This will be taken up below in §5.2.

5.2 Units, zeros and the factor N!

The various expressions Gibbs proposed as analogies for entropy, i.e. (77,81,84), were presented without any discussion of 'units and zeros', i.e. of their physical dimension and the constants that may be added to these expressions. This was only natural because Gibbs singled out those expressions for their formal merit of reproducing the fundamental equation, in which only the combination TdS appears. He discussed the question of the physical dimension of entropy by noting that the fundamental equation remains invariant if we multiply the analogue for temperature — i.e. the parameter θ in the canonical case, or the functions (80 or (83) for the microcanonical case — by some constant K and the corresponding analogues for entropy — (77), (81) and (84) — by $1/K$. Applied to the simple case of the monatomic ideal gas of N molecules, he concluded that, in order to equate the analogues of temperature to the ideal gas temperature, $1/K$ should be set equal to

$$(85)\quad \frac{1}{K} = \frac{2}{3}\frac{c_V}{N},$$

where c_V is the specific heat at constant volume. He notes that "this value had been recognized by physicists as a constant independent of the kind of monatomic gas considered" [Gibbs, 1902, p. 185]. Indeed, in modern notation, $1/K = k$, i.e. Boltzmann's constant.

Concerning the question of 'zeros', Gibbs noted that all the expressions proposed as analogy of entropy had the dimension of the logarithm of phase space volume and are thus affected by the choice of our units for length mass and time in the form of some additional constant (cf. [Gibbs, 1902, p. 19,183]). But even if some choice for such units is fixed, further constants could be added to the statistical analogs of entropy, i.e. arbitrary expressions that may depend on anything not varied in the fundamental equation. However, their values would disappear when differences of entropy are compared. And since only entropy differences have physical meaning, a question of determining these constants would thus appear to be immaterial. However, Gibbs went on to argue that "the principle that the entropy of any body has an arbitrary additive constant is subject to limitations when different quantities of the same substance are compared"[Gibbs, 1902, p. 206]. He formulated further conditions on how the additive constant may depend on the number N of particles in his final chapter.

Gibbs starts this investigation by raising the following problem. Consider the phase (i.e. microstate) $(\vec{q}_1, \vec{p}_1; \ldots; \vec{q}_N, \vec{p}_N)$ of an N-particle system where the particles are said to be "indistinguishable", "entirely similar" or "perfectly similar".[53] Now, if we perform a permutation on the particles of such a system, should we regard the result as a different phase or not? Gibbs first argues that it "seems in accordance with the spirit of the statistical method" to regard such phases as the

[53]Presumably, these terms mean (at least) that the Hamiltonian is invariant under their permutation, i.e. they have equal mass and interact in exactly the same way.

same. It might be urged, he says, that for such particles no identity is possible except that of qualities, and when comparing the permuted and unpermuted system, "nothing remains on which to base the identification of any particular particle of the first system with any particular particle of the second" [Gibbs, 1902, p. 187].

However, he immediately rejects this argument, stating that all this would be true for systems with "simultaneous objective existence", but hardly applies to the "creations of the imagination". On the contrary, Gibbs argues:

> "The perfect similarity of several particles of a system will not in the least interfere with the identification of a particular particle in one case and with a particular particle in another. The question is one to be decided in accordance with the requirements of practical convenience in the discussion of the problems with which we are engaged" [Gibbs, 1902, p. 188]

He continues therefore by exploring both options, calling the viewpoint in which permuted phases are regarded as identical the *generic* phase, and that in which they are seen as distinct the *specific* phase. In modern terms the generic phase space is obtained as the quotient space of the specific phase space obtained by identifying all phase points that differ by a permutation (see [Leinaas and Myrheim, 1977]). In general, there are $N!$ different permutations on the phase of a system of N particles,[54] and there are thus $N!$ different specific phases corresponding to one generic phase. This reduces the generic phase space measure by an overall factor of $\frac{1}{N!}$ in comparison to the specific phase space. Since the analogies to entropy all have a dimension equal to the logarithm of phase space measure, this factor shows up as an further additive constant to the entropy, namely $-\ln N!$ in comparison to an entropy calculated from the specific phase. Gibbs concludes that when N is constant, "it is therefore immaterial whether we use [the generic entropy] or [the specific entropy], since this only affects the arbitrary constant of integration which is added to the entropy [Gibbs, 1902, p. 206].[55]

However, Gibbs points out that this is *not* the case if we compare the entropies of systems with different number of particles. For example, consider two identical gases, each with the same energy U, volume V and number of particles N, in contiguous containers, and let the entropy of each gas be written as $S(U, V, N)$. Gibbs puts the entropy of the total system equal to the sum of the entropies:

$$(86)\quad S_{\text{tot}} = 2S(U, V, N).$$

Now suppose a valve is opened, making a connection between the two containers. Gibbs says that "we do not regard this as making any change in the entropy, although the gases diffuse into one another, and this process would increase the entropy if the gases were different" [Gibbs, 1902, p. 206-7]. Therefore, the entropy in this new situation is

[54]This assumes that the molecular states $\vec{p}_i, \vec{q}_i)$ of the particles do not coincide. However the points in specific phase space for which one or more molecular states do coincide constitute a set of Lebesgue measure zero anyway.

[55]The same conclusion also obtains for the Boltzmann entropy (61) [Huggett, 1999].

(87) $S'_{\rm tot} = S_{\rm tot}$.

But the new system, is a gas with energy $2U$, volume $2V$, and particle number $2N$. Therefore, we obtain:

(88) $S'_{\rm tot} = S(2U, 2V, 2N) = 2S(U, V, N)$,

where the right-hand side equation expresses the *extensivity* of entropy. This condition is satisfied (at least for large N) by the generic entropy but not by the specific entropy. Gibbs concludes "it is evident therefore that it is equilibrium with respect to generic phases, and not that with respect to specific, with which we have to do in the evaluation of entropy, ...except in the thermodynamics of bodies in which the number of molecules of the various kinds is constant" [Gibbs, 1902, p. 207].

The issue expressed in these final pages is perhaps the most controversial in Gibbs' book; at least it has generated much further discussion. Many later authors have argued that the insertion of a factor $1/N!$ in the phase space measure is obligatory to obtain "correct" results and, ultimately due to a lack of any metaphysical identity or "haecceity" of the perfectly similar particles considered. Some have even gone on to argue that quantum mechanics is needed to explain this. For example, [Huang, 1987, p. 154] writes "It is not possible to understand classically why we must divide [...] by $N!$ to obtain the correct counting of states. The reason is inherently quantum mechanical ...". However, many others deny this [Becker, 1967; van Kampen, 1984; Ray, 1984]. It would take me too far afield to discuss the various views and widespread confusion on this issue.

Let it suffice to note that Gibbs rejected arguments from the metaphysics of identity for the creations of the imagination. (I presume this may be taken to express that the phases of an N-particles system are theoretical constructs, rather than material objects.) Further, Gibbs did not claim that the generic view was correct and the specific view of incorrect; he preferred to settle the question by "practical convenience". There are indeed several aspects of his argument that rely on assumptions that may be argued to be conventional. for example the 'additivity' demand (86) could be expanded to read more fully:

(89) $S_{\rm tot}(U_1, V_1, N_1; U_2, V_2, N_2) + K_{\rm tot} = S_1(U_1, V_1, N_1) + K_1 + S_2(U_2, V_2, N_2) + K_2$,

Applied to the special case where S_1 and S_2 are identical functions taken at the same values of their arguments. The point to note here is that this relation only leads to (86) if we also employ the conventions $K_{\rm tot} = K_1 + K_2$ and $K_1 = K_2$. Also, his cautious choice of words concerning (87) — "we do not regard this as making any change" — suggest that he wants to leave open whether this equation expresses a fact or a conventional choice on our part. But by and large, it seems fair to say that Gibbs' criterion for practical convenience is simply the recovery of the properties usually assumed to hold for thermodynamic entropy.

As a final remark, note that the contrast mentioned here in passing by Gibbs, i.e. that in thermodynamics the mixing of identical gases, by allowing them to diffuse into one another, does not change the entropy, whereas this process does increase entropy if the gases are different, implicitly refers to an earlier discussion of this issue in his 1875 paper [Gibbs, 1906, pp. 165–167]. The contrast between the entropy of mixing of identical fluids and that of different fluids noted on that occasion is now commonly known as the *Gibbs paradox.* (More precisely, this 'paradox' is that the entropy of mixing different fluids is a constant ($kT \ln 2$ in the above case) as long as the substances are different, and vanishes abruptly when they are perfectly similar; thus negating the intuitive expectation one might have had that the entropy of mixing should diminish gradually when the substances become more and more alike). Now note that in the the specific view, mixing different substances and mixing identical substances both lead to an entropy increase: in that view there is no Gibbs paradox, since there is no abrupt change when the substances become more and more alike. On the other hand, the adoption of the generic view, i.e. the division of the phase space measure by $N!$, is used by Gibbs to recover the usual properties of thermodynamic entropy *including* the Gibbs paradox — the discontinuity between mixing of different and identical gases.

Still, many authors seem to believe that the division by $N!$ is a procedure that *solves* the Gibbs paradox. But this is clearly not the case; instead, it is the specific viewpoint that avoids the paradox, while the generic viewpoint recovers the Gibbs paradox for the statistical mechanical analogies to entropy. The irony of it all is that, in statistical mechanics, the term "Gibbs paradox" is sometimes used to mean or imply the *absence* of the original Gibbs paradox in the specific point of view, so that a resolution of *this* "Gibbs paradox" requires the return of the original paradox.

5.3 Gibbs on the increase of entropy

As we have seen, the Gibbs entropy may be defined as a functional on arbitrary probability density functions ρ on phase space Γ:[56]

$$(90)\quad \sigma[\rho] = -\int \rho(x) \ln \rho(x) dx$$

This expression has many well-known and useful properties. For example, under all probability densities restricted to the energy hypersurface $H(x) = E$, the microcanonical density (72) has the highest entropy. Similarly, one can show that of all distributions ρ with a given expectation value $\langle H \rangle_\rho$, the canonical distribution (73) has the highest entropy, and that of all distributions for which both $\langle H \rangle$ and $\langle N \rangle$ are given, the grand-canonical ensemble has the highest entropy.

[56] Gibbs actually does not use the term entropy for this expression. He calls the function $\ln \rho$ the "index of probability", and $-\sigma$ "the average index of probability". As we have seen, Gibbs proposed more than one candidate for entropy in the microcanonical ensemble, and was well aware that: " [t]here may be [...], and there are, other expressions that may be thought to have some claim to be regarded as the [...] entropy with respect to systems of a finite number of degrees of freedom" [Gibbs, 1902, p. 169].

But suppose that ρ is not stationary. It will therefore evolve in the course of time, as given by $\rho_t(x) = \rho(T_{-t}x)$. One might ask whether this entropy will increase in the course of time. However, Liouville's theorem implies immediately

(91) $\sigma[\rho_t] = \sigma[\rho_0]$.

In spite of the superficial similarity to Boltzmann's H, the Gibbs entropy thus remains constant in time. The explanation of the Second Law, or an approach to equilibrium, cannot be so simple.

However, Gibbs warns us to proceed with great caution. Liouville's theorem can be interpreted as stating that the motion of ρ_t can be likened to motion in a (multidimensional) incompressible fluid. He thus compared the evolution of ρ to that of the stirring of a dye in a incompressible medium [Gibbs, 1902, p. 143-151]. In this case too, the average density of the dye, as well as the average of any function of its density, does not change. Still, it is a familiar fact of experience that by stirring tends to bring about a uniform mixture, or a state with uniform density, for which the expression $-\int \rho \ln \rho \, dx$ would have increased to attain its maximum value.

Gibbs saw the resolution of this contradiction in the definition of the notion of density. This, of course, is commonly taken as the limit of the quantity of dye in a spatial volume element, when the latter goes to zero. If we apply this definition, i.e. take this limit first, and then consider the stirring motion, we will arrive at the conclusion that $-\int \rho \ln \rho dx$ remains constant. But if we consider the density defined for a fixed finite (non-zero) volume element, and then stir for an indefinitely long time, the density may become 'sensibly' uniform, a result which is not affected if we subsequently let the volume elements become vanishingly small. The problem, as Gibbs saw it, is therefore one of the order in which we proceed to take two limits.

Gibbs was aware that not all motions in phase space produce this tendency toward statistical equilibrium, just as not every motion in an incompressible fluid stirs a dy to a sensibly homogeneous mixture. Nevertheless, as he concluded tentatively,: "We might perhaps fairly infer from such considerations as have been adduced that an approach to a limiting condition of statistical equilibrium is the general rule, when the initial condition is not of that character" [Gibbs, 1902, p. 148].

5.4 Coarse graining

The most common modern elaboration of Gibbs' ideas is by taking recourse to a partitioning of phase space in cells, usually called "*coarse graining.* Instead of studying the original distribution function $\rho(x)$ we replace $\rho(x)dx$ by its phase average over each cell, by the mapping:

(92) $\mathcal{CG} : \rho(x) \mapsto \mathcal{CG}\rho(x) = \sum_i \hat{\rho}(i) \mathbb{1}_{\omega_i}(x)$,

where

$$(93)\quad \hat{\rho}(i) := \frac{\int_{\omega_i} \rho(x)dx}{\int_{\omega_i} dx},$$

and $\mathbb{1}$ denotes the characteristic function:

$$(94)\quad \mathbb{1}_A(x) = \begin{cases} 1 & \text{if } x \in A, \\ 0 & \text{elsewhere.} \end{cases}$$

The usual idea is that such a partition matches the finite precision of our observational capabilities, so that a "coarse grained" distribution might be taken as a sufficient description of what is observable. Obviously, the average value of any function on Γ that does not vary too much within cells is approximately the same, whether we use the fine-grained or the coarse-grained distribution.

For any ρ one can also define the coarse grained entropy $\Sigma[\rho]$ as the composition of (92) and (90):

$$(95)\quad \Sigma[\rho] := \sigma[\mathcal{CG}\rho].$$

This coarse grained entropy need not be conserved in time. Indeed, it is easy to show (cf. [Tolman, 1938, p. 172]). that:

$$(96)\quad \Sigma[\rho] \geq \sigma[\rho].$$

Hence, if we assume that at some initial time that $\rho_0 = \mathcal{CG}\rho_0$, e.g. if $\rho_0 \propto \frac{1}{V_i}\mathbb{1}_{\omega_i}$ for some cell i, then for all t:

$$(97)\quad \Sigma[\rho_t] \geq \sigma[\rho_t] = \sigma[\rho_0] = \Sigma[\rho_0].$$

However, this does not imply that $\Sigma[\rho_t]$ is non-decreasing or that it approaches a limiting value as $t \longrightarrow \infty$.

If a property, similar to the stirring of a dye holds for the dynamical evolution of ρ_t, one may have

$$(98)\quad \lim_{t\longrightarrow\infty} \Sigma[\rho_t] = \Sigma[\rho_{\mathrm{mc}}]$$

and hence, an approach towards equilibrium could emerge on the coarse-grained level. This convergence will of course demand a non-trivial assumption about the dynamics. In modern work this assumption is that the system has the *mixing* property (see paragraph 6.1).

5.5 Comments

Gibbs' statistical mechanics has produced a formalism with clearly delineated concepts and methods, using only Hamiltonian mechanics and probability theory. It can and is routinely used to calculate equilibrium properties of gases and other systems by introducing a specific form of the Hamiltonian. The main problems that Gibbs has left open are, first, the motivation for the special choice of the equilibrium ensembles and, second, that the quantities serving as thermodynamic analogies are not uniquely defined. However, much careful work has been devoted to show that, under certain assumptions about tempered interaction of molecules, unique thermodynamic state functions, with their desired properties are obtained in the 'thermodynamic limit' (cf. §6.3).

1. Motivating the choice of ensemble. While Gibbs had not much more to offer in recommendation of these three ensembles than their simplicity as candidates for representation for equilibrium, modern views often provide an additional story. First, the microcanonical ensemble is particularly singled out for describing an ensemble of systems in thermal isolation with a fixed energy E.

Arguments for this purpose come in different kinds. As argued by Boltzmann (1868), and shown more clearly by Einstein (1902), the microcanonical ensemble is the *unique* stationary density for an isolated ensemble of systems with fixed energy, if one assumes the ergodic hypothesis. Unfortunately, for this argument, the ergodic hypothesis is false for any system that has a phase space of dimension 2 or higher (cf. paragraph 6.1).

A related but more promising argument relies on the theorem that the measure $P_{\rm mc}$ associated with the microcanonical ensemble via $P_{\rm mc}(A) = \int_A \rho_{\rm mc}(x)dx$ is the unique stationary measure under all measures that are absolutely continuous with respect to $P_{\rm mc}$, if one assumes that the system is metrically transitive (again, see paragraph 6.1).

This argument is applicable for more general systems, but its conclusion is weaker. In particular, one would now have to argue that physically interesting systems are indeed metrically transitive, and why measures that are not absolutely continuous with respect to the microcanonical one are somehow to be disregarded. The first problem is still an open question, even for the hard-spheres model (as we shall see in paragraph 6.1). The second question can be answered in a variety of ways.

For example, [Penrose, 1979, p. 1941] adopts a principle that every ensemble should be representable by a (piecewise) continuous density function, in order to rule out "physically unreasonable cases". (This postulate implies absolute continuity of the ensemble measure with respect to the microcanonical measure by virtue of the Radon-Nikodym theorem.) See [Kurth, 1960, p. 78] for a similar postulate. Another argument, proposed by [Malament and Zabell, 1980], assumes that the measure P associated with a physically meaningful ensemble should have a property called 'translation continuity'. Roughly, this notion means that the probability assigned to any measurable set should be a continuous function under small displacements of that set within the energy hypersurface. Malament & Zabell show that this property is equivalent to absolute continuity of P with respect to $\mu_{\rm mc}$, and thus singles out the microcanonical measure uniquely if the system is metrically transitive (see [van Lith, 2001b, for a more extensive discussion]).

A third approach, due to Tolman and Jaynes, more or less postulates the microcanonical density, as a appropriate description of our knowledge about the microstate of a system with given energy (regardless of whether the system is metrically transitive or not).

Once the microcanonical ensemble is in place as a privileged description of an isolated system with a fixed energy, one can motivate the corresponding status for the other ensembles with relatively less effort. The canonical distribution is shown to provide the description of a small system S_1 in weak energetic contact with a larger system S_2, acting as a 'heat bath' (see [Gibbs, 1902, p. 180–183]). Here, it is assumed that the total system is isolated and described by a microcanonical ensemble, where the total system has a Hamiltonian $H_{\rm tot} = H_1 + H_2 + H_{\rm int}$ with $H_2 \gg H_1 \gg H_{\rm int}$. More elaborate versions of such an argument are given by Einstein (1902) and Martin-Löf (1979). Similarly, the grand-canonical ensemble can be derived for a small system that can exchange both energy and particles with a large system. (see [van Kampen, 1984]).

2. The 'equivalence' of ensembles. It is often argued in physics textbooks that the choice between these different ensembles (say the canonical and microcanonical) is deprived of practical relevance by a claim that they are all "equivalent". (See [Lorentz, 1916, p. 32] for perhaps the earliest version of this argument, or [Thompson, 1972, p. 72; Huang, 1987, p. 161-2] for recent statements.) What is meant by this claim is that if the number of constituents increases, $N \longrightarrow \infty$, and the total Hamiltonian is proportional to N, the thermodynamic relations derived from each of them will coincide in this limit.

However, these arguments should not be mistaken as settling the *empirical* equivalence of the various ensembles, even in this limit. For example, it can be shown that the microcanonical ensemble admits the description of certain metastable thermodynamic states, (e.g. with negative heat capacity) that are excluded in the canonical ensemble (see [Touchette, 2003; Touchette et al., 2004, and literature cited therein]).

3. The coarse-grained entropy. The coarse-graining approach is reminiscent of Boltzmann's construction of cells in his (1877b); cf. the discussion in paragraph 4.4). The main difference is that here one assumes a partition on phase-space Γ, where Boltzmann adopted it in the μ-space. Nevertheless, the same issues about the origin or status of a privileged partition can be debated (cf. p. 977). If one assumes that the partition is intended to represent what we *know* about the system, i.e. if one argues that all we know is whether its state falls in a particular cell ω_i, it can be argued that the its status is subjective. If one argues that the partition is meant to represent limitations in the precision of human observational possibilities, perhaps enriched by instruments, i.e. that we cannot observe more about the system than that its state is in some cell ω_i, one might argue that its choice is objective, in the sense that there are objective facts about what a given epistemic community can observe or not. Of course, one can then still maintain that the status of the coarse-graining would then be anthropocentric (see also the discussion in §7.5). However, note that Gibbs himself did not argue for a preferential size of the cells in phase space, but for taking the limit in which their size goes to zero in a different order.

4. Statistical equilibrium. Finally, a remark about Gibbs' notion of equilibrium. This is fundamentally different from Boltzmann's 1877 notion of equilibrium as the macrostate corresponding to the region occupying the largest volume in phase space (cf. section 4.4). For Gibbs, statistical equilibrium can only apply to an ensemble. And since any given system can be regarded as belonging to an infinity of different ensembles, it makes no sense to say whether an individual system is in statistical equilibrium or not. In contrast, in Boltzmann's case, equilibrium can be attributed to a single system (namely if the microstate of that system is an element of the set $\Gamma_{\rm eq} \subset \Gamma$). But it is not guaranteed to remain there for all times.

Thus, one might say that in comparison with the orthodox thermodynamical notion of equilibrium (which is both stationary and a property of an individual system) Boltzmann (1877b) and Gibbs each made an opposite choice about which aspect to preserve and which aspect to sacrifice. See [Uffink, 1996b; Callender, 1999; Lavis, 2005] for further discussions.

6 MODERN APPROACHES TO STATISTICAL MECHANICS

This section will leave the more or less historical account followed in the previous sections behind, and present a selective overview of some influential modern approaches to statistical physics. In particular, we focus on ergodic theory (§ 6.1–6.2), the theory of the thermodynamic limit §6.3, the work of Lanford on the Boltzmann equation (§6.4), and the BBGKY approach in §6.5.

6.1 Ergodic theory

When the Ehrenfests critically reviewed Boltzmann's and Gibbs' approach to statistical physics in their renowned Encyclopedia article 1912, they identified three issues related to the ergodic hypothesis.

1. The ambiguity in Boltzmann's usage of "probability" of a phase space region (as either the relative volume of the region or the relative time spent in the region by the trajectory of the system).

2. The privileged status of the microcanonical probability distribution or other probability distributions that depend only on the Hamiltonian.

3. Boltzmann's argument that the microstate of a system, initially prepared in a region of phase space corresponding to a non-equilibrium macrostate, should tend to evolve in such a way that its trajectory will spend an overwhelmingly large majority of its time inside the region of phase space corresponding to the equilibrium macrostate $\Gamma_{\rm eq}$.

In all these three problems, a more or less definite solution is obtained by adopting the ergodic hypothesis. Thus, the Ehrenfests suggested that Boltzmann's answer to the above problems *depended* on the ergodic hypothesis. As we have seen, this is correct only for Boltzmann's treatment of issue (2) in his 1868a. The doubtful status of the ergodic hypothesis, of course, highlighted the unresolved status of these problems in the Ehrenfests' point of view.

In later works the "ergodic problem" has become more exclusively associated with the first issue on the list above, i.e., the problem of showing the equality of phase and time averages. This problem can be formulated as follows. Consider a Hamiltonian system and some function f defined on its phase space Γ. The (infinite) time average of f, for a system with initial state x_0 may be defined as:

(99) $$\overline{f(x_0)} = \lim_{T\to\infty} \frac{1}{T} \int_0^T f(T_t x_0) dt$$

where T_t is the evolution operator. On the other hand, for an ensemble of systems with density $\rho_t(x)$, the ensemble average of f is

(100) $$\langle f \rangle_t = \int f(x)\rho_t(x) dx.$$

The ergodic problem is the question whether, or under which circumstances, the time average and ensemble average are equal, i.e.: $\overline{f(x_0)} \stackrel{?}{=} \langle f \rangle_t$. Note that there are immediate differences between these averages. $\overline{f}$ depends on the initial state x_0, in contrast to $\langle f \rangle$. Indeed, each choice of an initial phase point gives rise to another trajectory in phase space, and thus gives, in general, another time average. Secondly, $\langle f \rangle$ will in general depend on time, whereas $\overline{f}$ is time-independent. Hence, a general affirmative answer to the problem cannot be expected.

However, in the case of a stationary ensemble (statistical equilibrium) the last disanalogy disappears. Choosing an even more special case, the microcanonical ensemble $\rho_{\rm mc}$, the simplest version of the ergodic problem is the question:

(101) $$\overline{f(x_0)} \stackrel{?}{=} \langle f \rangle_{\rm mc}.$$

Now it is obvious that if Boltzmann's ergodic hypothesis is true, i.e. if the trajectory of the system traverses all points on the energy hypersurface Γ_E, the desired equality holds. Indeed, take two arbitrary points x and y in Γ_E. The ergodic hypothesis implies that there is a time τ such that $y = T_\tau x$. Hence:

$$\begin{aligned}
\overline{f(y)} &= \lim_{T\to\infty} \frac{1}{T} \int_0^T f(T_{t+\tau} x) dt \\
&= \lim_{T\to\infty} \frac{1}{T} \left(\int_0^\tau f(T_t x) dt + \int_0^T f(T_t x) dt \right) \\
&= \lim_{T\to\infty} \frac{1}{T} \int_0^T f(T_t x) dt = \overline{f(x)}
\end{aligned}$$

In other words, $\overline{f}$ must be constant over Γ_E, and hence, also equal to the microcanonical expectation value.

For later reference we note another corollary: the ergodic hypothesis implies that $\rho_{\rm mc}$ is the only stationary density on Γ_E (cf. section 4.1).

The Ehrenfests doubted the validity of the ergodic hypothesis, as Boltzmann had himself, and therefore proposed an alternative, which they called the *quasi-ergodic hypothesis*. This states that the trajectory lies dense in Γ_E, i.e., x_t will pass through every open subset in Γ_E, and thus come arbitrarily close to every point in Γ_E. The system may be called quasi-ergodic if this holds for all its trajectories. As we have seen, this formulation seems actually closer to what Boltzmann may have intended, at least in 1871, than his own literal formulation of the hypothesis.

Not long after the Ehrenfests' review, the mathematical proof was delivered that the ergodic hypothesis cannot hold if Γ_E is a more than one-dimensional manifold [Rosenthal, 1913; Plancherel, 1913]. The quasi-ergodic hypothesis, on the other hand, cannot be immediately dismissed. In fact, it may very well be satisfied for Hamiltonian systems of interest to statistical mechanics. Unfortunately, it has remained unclear how it may contribute to a solution to the ergodic problem. One might hope, at first sight, that for a quasi-ergodic system time averages and microcanonical averages coincide for continuous functions, and that the microcanonical density $\rho_{\rm mc}$ is the only continuous stationary density. But even this is unknown. It is known that quasi-ergodic systems may fail to have a unique stationary measure [Nemytskii and Stepanov, 1960, p. 392]. This is not to say that quasi-ergodicity has remained a completely infertile notion. In topological ergodic theory, the condition is known under the name of "minimality", and implies several interesting theorems (see [Petersen, 1983, p. 152ff]).

While the Rosenthal-Plancherel result seemed to toll an early death knell over ergodic theory in 1913, a unexpected revival occurred in the early 1930s. These new results were made possible by the stormy developments in mathematics and triggered by Koopman's results, showing how Hamiltonian dynamics might be embedded in a Hilbert space formalism where the evolution operators T_t are represented as a unitary group. This made a whole array of mathematical techniques (e.g. spectral analysis) available for a new attack on the problem.

The first result was obtained by von Neumann in a paper under the promising (but misleading) title "Proof of the Quasi-Ergodic Hypothesis" 1932. His theorem was strengthened by G.D. Birkhoff in a paper entitled "Proof of the Ergodic Theorem" 1931, and published even before von Neumann's.

Since their work, and all later work in ergodic theory, involves more precise mathematical notions, it may be worthwhile first to introduce a more abstract setting of the problem. An abstract *dynamical system* is defined as a tuple $\langle \Gamma, \mathcal{A}, \mu, T \rangle$, where Γ as an arbitrary set, $\mathcal{A}$ is a σ-algebra of subsets of Γ, called the 'measurable' sets in Γ, and μ is a probability measure on Γ, and T denotes a one-parameter group of one-to-one transformations T_t on Γ (with $t \in \mathbb{R}$ or $t \in \mathbb{Z}$) that represent the evolution operators. The transformations T_t are assumed to be measure-preserving, i.e. $\mu(T_t A) = \mu(A)$ for all $A \in \mathcal{A}$. In the more concrete setting of statistical mechanics, one may take Γ to be the energy hypersurface, $\mathcal{A}$ the collection of its Borel subsets, μ the microcanonical probability measure and T the evolution induced by the Hamiltonian equations.

The von Neumann-Birkhoff ergodic theorem can be formulated as follows:

> ERGODIC THEOREM: Let $\langle \Gamma, \mathcal{A}, \mu, T \rangle$ be any dynamical system and f be an integrable function on Γ. Then
>
> (i) $\overline{f(x)} = \lim_{T\to\infty} \frac{1}{T}\int_0^T f(T_t x)dt$ exists for almost all x; i.e. the set of states $x \in \Gamma$ for which $\overline{f(x)}$ does not exist has μ-measure zero.
>
> (ii) $\overline{f(x)} = \langle f \rangle_\mu$ for almost all x iff the system is metrically transitive.

Here, metric transitivity means that it is impossible is to carve up Γ in two regions of positive measure such that any trajectory starting in one region never crosses into the other. More precisely:

> METRIC TRANSITIVITY: A dynamical system is called metrically transitive[57] iff the following holds: for any partition of Γ into disjoint sets A_1, A_2 such that $T_t A_1 = A_1$ and $T_t A_2 = A_2$, it holds that $\mu(A_1) = 0$ or $\mu(A_2) = 0$.

It is not difficult to see why this theorem may be thought of as a successful solution of the original ergodic problem under a slight reinterpretation. First, metric transitivity captures in a measure-theoretic sense the idea that trajectories wander wildly across the energy hypersurface, allowing only exceptions for a measure zero set. Secondly, the theorem ensures the equality of time and microcanonical ensemble average, although only for integrable functions and, again, with the exception of a measure zero set. But that seemed good enough for the taste of most physicists.

The ergodic theorem was therefore celebrated as a major victory. In the words of Reichenbach:

> Boltzmann introduced [...] under the name of *ergodic hypothesis* [...] the hypothesis that the phase point passes through every point of the energy hypersurface. This formulation is easily shown to be untenable. It was replaced by P. and T. Ehrenfest by the formulation that the path comes close to every point within any small distance ϵ which we select and which is greater than 0.
>
> There still remained the question whether the ergodic hypothesis must be regarded as an independent supposition or whether it is derivable from the canonical equations, as Liouville's theorem is.

[57]This name is somewhat unfortunate, since the condition has nothing to do with metric in the sense of distance, but is purely measure-theoretical. Metrically transitive systems are also called 'metrically indecomposable' or, especially in the later literature 'ergodic'. I will stick to the older name in order to avoid confusion with the ergodic hypothesis.

> This problem[...] was finally solved through ingenious investigations by John von Neumann and George Birkhoff, who were able to show that the second alternative is true. [...] With von Neumann and Birkhoff's theorem, deterministic physics has reached its highest degree of perfection: the strict determinism of elementary processes is shown to lead to statistical laws for macroscopic occurrences." [Reichenbach, 1956, p. 78]

Unfortunately, nearly everything stated in this quotation is untrue.

Problems

1. Do metrically transitive systems exist? An immediate question is of course whether metrically transitive systems exist. In a mathematical sense of 'exist' the answer is affirmative. More interesting is the question of whether one can show metric transitivity for any model that is realistic enough to be relevant to statistical mechanics.

A few mechanical systems have been explicitly proven to be metrically transitive. For example: one hard sphere moving in a vessel with a convex scatterer, or a disc confined to move in a 'stadium' (two parallel line-segments connected by two half circles) or its three-dimensional analogue: one hard sphere moving in a cylinder, closed on both sides by half-spheres. But in statistical mechanics one is interested in systems with many particles.

In [1963], Sinai announced he had found a proof that a gas consisting of N hard spheres is metrically transitive. The ergodic theorem thus finally seemed to be relevant to physically interesting gas models. Of course, the hard-spheres-model is an idealization too, but the general expectation among physicists was that a transition to more sophisticated models of a gas system would only make the metric transitivity even more likely and plausible, even though admittedly harder to prove.

The problem proves to be extraordinarily tedious, and Sinai's proof was complicated and, actually, never completely published. But many partial results were. In fact, the development of ideas and techniques needed for the effort contributed much to the emergence of a vigorous mathematical theory, nowadays called 'ergodic theory'. And since Sinai's claim seemed so desirable, many books and articles presented the claim as a solid proven fact (e.g. [Lebowitz and Penrose, 1973; Sklar, 1993]).

But by the 1980s, the delay in the publication of a complete proof started to foster some doubts about the validity of the claim. Finally, [Sinai and Chernov, 1987, p. 185] wrote: "The announcement made in [[Sinai, 1963]] for the general case must be regarded as immature." What has been shown rigorously is that a system of three hard spheres is metrically transitive. Recently, the problem has been taken further by [Szász, 1996] and [Simányi and Szász, 1999]. They have ascertained that for a model of N hard spheres, the ergodic component, i.e. a subset of the energy hypersurface on which the assumption of metric transitivity holds has positive measure. The full problem, however, still awaits solution.

2. Infinite times. In the definition of the time average (99) the limit $T \to \infty$ is taken. This brings along a number of problems:

(i). The time average is interesting because it is experimentally accessible. The hope is that it represents the equilibrium value of f. But the limit $T \to \infty$ tells us nothing about what happens in a finite time. What is empirically accessible, at best, is the quantity $\frac{1}{T}\int_0^T f(T_t x_0)dt$ for a large but finite T. This expression can still deviate arbitrarily far from the limiting value.

(ii). The limit may even exist while the system is not in equilibrium. A time-averaged value need not be an equilibrium value, because in general

$$\text{(102)} \quad \lim_{T\to\infty} \frac{1}{T}\int_0^T f(T_t x)\, dt \neq \lim_{t\to\infty} f(T_t x).$$

For periodical motions, for example, the left-hand side exists but the right-hand side does not.

(iii) Empirically, equilibrium often sets in quite rapidly. But the time T needed to make $\frac{1}{T}\int_0^T f(T_t x_0)dt$ even remotely close to $\langle f\rangle_{\rm mc}$ might be enormous, namely of the order of Boltzmann's estimate of the Poincaré-recurrence times! (See also [Jaynes, 1967, p. 94].)

3. The measure-zero problem. The result that the ergodic theorem provides is that for metrically transitive systems $\overline{f(x)} = \langle f\rangle_{\rm mc}$ except for a set of microstates with measure zero. So the suggestion here is that this set of exceptions is in some sense negligible. And, as judged from the probability measure $\mu_{\rm mc}$, that is obviously true. But a set of measure zero need not be negligible in any other sense. It is well-known that if one compares 'smallness in measure' with other natural criteria by which one can judge the 'size' of sets, e.g. by their cardinality, dimension or Baire category, the comparisons do not match. Sets of measure zero can be surprisingly large by many other standards [Sklar, 1993, pp. 181–188].

More importantly, one might choose another measure μ', such that μ-measure zero sets are no longer sets of μ'-measure zero and conversely. It is of course the choice of the measure that determines which sets have measure zero. Thus, if one decides to disregard or neglect sets with a microcanonical measure zero, a privileged status of the microcanonical measure is already presupposed. But this means the virtue of the ergodic theorem as a means of motivating a privileged role of the microcanonical measure is diminished to a self-congratulating one.

6.2 *The mixing property, K systems and Bernoulli systems*

Ergodic theory, the mathematical field that emerged from the theorems of Birkhoff and von Neumann, may be characterized as a rigorous exploration of the question to what extent a deterministic, time-reversal invariant dynamical system may give rise to random-like behaviour on a macroscopic scale, by assuming various special properties on its dynamics.

In its modern incarnation, this theory distinguishes a hierarchy of such properties that consists of various strengthenings of metric transitivity. Perhaps the most important are the mixing property, the property of being a 'K system' and the Bernoulli systems. The higher up one goes this ladder, the more 'random' behaviour is displayed. The evolution at the microlevel is in all cases provided by the deterministic evolution laws. In the (extensive) literature on the subject, many more steps in the hierarchy are distinguished (such as 'weak mixing', 'weak Bernoulli', 'very weak Bernoulli' etc.), and also some properties that do not fit into a strict linear hierarchy (like the 'Anosov' property, which relies on topological notions rather than on a purely measure-theoretical characterization of dynamical systems). It falls beyond the scope of this paper to discuss them.

Mixing

The idea of *mixing* is usually attributed to Gibbs, in his comparison of the evolution of ensembles with stirring of a dye into an incompressible fluid (cf. section 5.4). Even if initially the fluid and the dye particles occupy separate volumes, stirring will eventually distribute the dye particles homogeneously over the fluid. The formal definition is:

> MIXING: A dynamical system $\langle \Gamma, \mathcal{A}, \mu, T \rangle$ is called mixing iff $\forall A, B \in \mathcal{A}$
>
> (103) $\lim_{t \to \infty} \mu(T_t A \cap B) = \mu(A)\mu(B)$.

In an intuitive sense the mixing property expresses the idea that the dynamical evolution will thoroughly stir the phase points in such a way that points initially contained in A eventually become homogeneously distributed over all measurable subsets B of Γ. One can easily show that *mixing* is indeed a stronger property than metric transitivity, by applying the condition to an invariant set A and choosing $B = A$. The converse statement does not hold. (E.g.: the one-dimensional harmonic oscillator is metrically transitive but not mixing).

Again, there is an interesting corollary in terms of probability measures or densities. Consider a mixing system, and a time-dependent probability density ρ_t, such that ρ_t is *absolutely continuous* with respect to the microcanonical measure μ. (This means that all sets $A \in \mathcal{A}$ with $\mu(A) = 0$, also have $\int_A \rho_t(x)dx = 0$, or equivalently, that ρ_t is a proper density function that is integrable with respect to μ.) In this case, the probability measure associated with ρ_t converges, as $t \longrightarrow \infty$, to the microcanonical measure. Thus, an ensemble of mixing systems with an absolutely continuous density will asymptotically approach to statistical equilibrium. Note that the same result will also hold for $t \longrightarrow -\infty$, so that there is no conflict with the time reversal invariance. Is it in conflict with Poincaré's recurrence theorem? No, the recurrence theorem is concerned with microstates (phase points), and not probability densities. Even when almost all trajectories eventually return close by their original starting point, the recurrence time will differ for each phase point, so that the evolution of an ensemble of such points can show a definite approach to statistical equilibrium.

Note also that if the result were used as an argument for the privileged status of the microcanonical measure (viz., as the unique measure that all absolutely continuous probability distributions evolve towards), the strategy would again be marred by the point that the condition of absolute continuity already refers to the microcanonical measure as a privileged choice.

Despite the elegance of the mixing property, we can more or less repeat the critical remarks made in the context of the ergodic theorem. In the first place, the condition considers the limit $t \to \infty$, which implies nothing about the rate at which convergence takes place. Secondly, the condition imposed is trivially true if we choose A or B to be sets of measure zero. Thus, the mixing property says nothing about the behaviour of such sets during time evolution. And thirdly, one is still faced with the question whether the mixing property holds for systems that are physically relevant for statistical mechanics. And since the property is strictly stronger than metric transitivity, this problem is at least as hard.

K systems

The next important concept is that of a *K system* ('K' after Kolmogorov). For simplicity, we assume that time is discrete, such that $T_t = T^t$, for $t \in \mathbb{Z}$. There is a perfectly analogously defined concept for continuous time, called K flows (cf. Emch, this volume, Definition 10.3.2).

> K SYSTEM:[58] A dynamical system $\langle \Gamma, \mathcal{A}, \mu, T \rangle$ is called a *K system* if there is a subalgebra $\mathcal{A}_0 \subset \mathcal{A}$, such that
>
> 1. $T^n\mathcal{A}_0 \subset T^m\mathcal{A}_0$ for times $m < n$; where $\subset$ denotes proper inclusion.
> 2. the smallest σ-algebra containing $\cup_{n=1}^{\infty} T^{-n}\mathcal{A}_0$ is $\mathcal{A}$.
> 3. $\cap_{n=1}^{\infty} T^n\mathcal{A}_0 = \mathcal{N}$, where $\mathcal{N}$ is the σ-algebra containing only sets of μ-measure zero or one.

At first sight, this definition may appear forbiddingly abstract. One may gain some intuition by means of the following example. Consider a finite partition

[58]There is a considerable variation in the formulation of this definition [Cornfeld *et al.*, 1982; Batterman, 1991; Berkovitz et al., 2006]. The present formulation adds one more. It is identical to more common definitions if one replaces n and m in the exponents of T by $-n$ and $-m$ respectively.

$\alpha = \{A_1, \ldots A_m\}$ of Γ into disjoint cells and the so-called *coarse-grained history* of the state of the system with respect to that partition. That is, instead of the detailed trajectory x_t, we only keep a record of the labels i of the cell A_i in which the state is located at each instant of time, until time t=0:

$$(104)\quad \ldots i_{-k}, \ldots, i_{-3}, i_{-2}, i_{-1}, i_0 \qquad i_{-k} \in \{1, \ldots, m\}, \quad k \in \mathbb{N}.$$

This sequence is completely determined by the microstate x at $t = 0$:

$$(105)\quad i_{-k}(x) = \sum_{j=1}^{m} j \mathbb{1}_{A_j}(T^{-k}x)$$

where $\mathbb{1}$ denotes the characteristic function (94). Yet, as we shall see, for a K system, this sequence typically behaves in certain respects like a random sequence. Observe that

$$(106)\quad i_{-k}(x) = j \Longleftrightarrow T^{-k}x \in A_j \Longleftrightarrow x \in T^k A_j;$$

so we can alternatively express the coarse-grained history by means of evolutions applied to the cells in the partition. If $T\alpha := \{TA_1, \ldots, TA_m\}$, let $\alpha \vee T\alpha := \{A_i \cup TA_j \ : i, j = 1, \ldots m\}$ denote the common refinement of α and $T\alpha$. Saying that x belongs to $A_i \cup TA_j$ is, of course, equivalent to providing the last two terms of the sequence (104). Continuing in this fashion, one can build the refinement

$$(107)\quad \bigvee_{k=0}^{\infty} T^k\alpha = \alpha \vee T\alpha \vee T^2\alpha \cdots \vee T^k\alpha \vee \cdots,$$

each element of which corresponds to a particular coarse-grained history (104) up to t=0. The collection (107) is no longer finite, but still a countable partition of Γ.

Now take $\mathcal{A}_0$ to be the σ-algebra generated from the partition $\bigvee_{k=0}^{\infty} T^k\alpha$. Clearly, the events in this algebra are just those whose occurrence is completely decided whenever the coarse-grained history is known. In other words, for all $A \in \mathcal{A}_0$, $\mu(A|C)$ is zero or one, if C is a member of (107). It is easy to see that $T^{-m}\mathcal{A}_0$ is just the σ-algebra generated from $T^{-m} \bigvee_{k=0}^{\infty} T^k\alpha = \bigvee_{k=-m}^{\infty} T^k\alpha$, i.e. from the partition characterizing the coarse-grained histories up to $t = m$. Since the latter partition contains the history up to $t = n$ for all $n < m$, we have:

$$(108)\quad T^{-m}\mathcal{A}_0 \subseteq T^{-n}\mathcal{A}_0 \text{ for all } n < m.$$

This is equivalent to condition 1, but with '$\subset$' replaced by '$\subseteq$'.

Further, to explain condition 2, note that the smallest σ-algebra containing $\cup_{n=1}^{N} T^{-n}\mathcal{A}_0$ is generated by the union of the partitions $\bigvee_{k=-n}^{\infty} T^k\alpha$ for all $n \leq N$, which in view of (108) is just $T^{-N}\mathcal{A}_0$. Thus, condition 2 just says that if we extend the record of the coarse-grained history to later times $t = N > 0$, and let $N \longrightarrow \infty$, the partition eventually becomes sufficiently fine to generate all measurable sets in $\mathcal{A}$. This is a strong property of the dynamics. It means that the entire coarse-grained record, extending from $-\infty$ to ∞, provides all information needed to separate all the measurable sets in $\mathcal{A}$, (except, possibly, if they differ by a measure zero set.)

Similarly, in order to explain condition 3, note that (108) implies that $\cap_{n=1}^{N} T^n \mathcal{A}_0 = T^N \mathcal{A}_0$, which is generated from $\bigvee_{k=0}^{\infty} T^k \alpha$, i.e., the coarse-grained histories up to time $-N$. Thus, condition 3 expresses the demand that, as we let $N \longrightarrow \infty$, the class of events that are settled by the coarse-grained histories up to time $t = -N$ shrinks to the 'trivial' algebra of those sets that have probability one or zero. In other words, for every event $A \in \mathcal{A}$, with $0 < \mu(A) < 1$, the occurrence of A is undecided at some early stage of the coarse-grained history.

Yet the truly remarkable feature of K systems lies in the strict inclusion demanded in condition 1: at any time n, the collection of events decided by the coarse-grained histories up to n, is strictly smaller than the collection of events decided at time $n+1$. Since the latter is generated from the former by adding the partition $T^{-(n+1)}\alpha$ to the partition $\bigvee T^{-k}\alpha$, this means that at each time n the question which cell of the partition is occupied at time $n+1$ is not answerable from the knowledge of the previous coarse-grained history. This is quite a remarkable property for a sequence generated by a deterministic law of motion, although, of course, it is familiar for random sequences such as tosses with a die or spins of a roulette wheel.

In this attempt at elucidation, we have presupposed a particular finite partition α. One may ask whether there always is, for each Kolmogorov system, such a partition. The answer is yes, provided the system obeys some mild condition (that $\langle \Gamma, \mathcal{A}, \mu \rangle$ is a *Lebesgue space*).[59] Another question is whether the claims made about coarse-grained histories are specific for this particular partition. The answer is no. One may show that, given that they hold for some partition α, they also hold for *any* choice of a finite partition of Γ. (Very roughly speaking: because the partition $\bigvee_n T^n \alpha$ generates the σ-algebra of all events, the coarse-grained histories constructed from another finite partition can be reconstructed in terms of the coarse-grained histories in terms of α.

Bernoulli systems

The strongest property distinguished in the ergodic hierarchy is that of *Bernoulli systems*. To introduce the definition of this type of dynamical systems, it is useful to consider first what is usually known as a 'Bernoulli' scheme. Consider an elementary chance set-up with outcomes $\{A_1, \ldots A_m\}$ and probabilities p_j. A Bernoulli scheme is defined as the probability space obtained from doubly infinite sequences of independent identically distributed repetitions of trials on this elementary set-up. Formally, a Bernoulli scheme for a set (or "alphabet") $\alpha = \{1, \ldots, m\}$ with

[59]Roughly, this condition means that $\langle \Gamma, \mathcal{A}, \mu \rangle$ is isomorphic (in a measure-theoretic sense) to the interval $[0, 1]$, equipped with the Lebesgue measure. (See [Cornfeld *et al.*, 1982, p. 449] for the precise definition).

probabilities $\{p_j\}$ is the probability space $\langle \Gamma, \mathcal{A}, \mu \rangle$, where Γ is the set of all doubly infinite sequences

(109) $\eta = (\ldots, i_{-2}, i_{-1}, i_o, i_1, i_2 \ldots,) \qquad i_k \in \{1, \ldots, m\}; k \in \mathbb{Z}$

and $\mathcal{A}$ is defined as the smallest σ-algebra on Γ containing the sets:

(110) $A_k^j := \{\eta \in \Gamma \ : \ i_k = j\}.$

$\mathcal{A}$ is also known as the cylinder algebra. Further, we require of a Bernoulli scheme that:

(111) $\mu(A_k^j) = p_j$ for all $k \in \mathbb{Z}$.

One can turn this probability space into a dynamical system by introducing the discrete group of transformations T^m, $m \in \mathbb{Z}$, where T denotes the shift, i.e. the transformation on Γ that shifts each element of a sequence η one place to the left:

(112) For all $k \in \mathbb{Z}$: $\quad T(i_k) = i_{k-1}$.

Thus we define:

> BERNOULLI SYSTEM: A dynamical system $\langle \Gamma, \mathcal{A}, \mu, T \rangle$ with a discrete time evolution T is a Bernoulli-system iff there is a finite partition $\alpha = \{A_1, \ldots, A_m\}$ of Γ such that the doubly infinite coarse-grained histories are (isomorphic to) a Bernoulli scheme for α with distribution
>
> (113) $p_i = \mu(A_i) \quad i \in \{1, \ldots m\}$.

Thus, for a Bernoulli system, the coarse-grained histories on α behave as randomly as independent drawings from an urn. These histories show no correlation at all, and the best prediction one can make about the location of the state at time $n+1$, even if we know the entire coarse-grained history from minus infinity to time n, is no better than if we did not know anything at all. One can show that every Bernoulli-system is also a K-system, but that the converse need not hold.

Discussion

Ergodic theory has developed into a full-fledged mathematical discipline with numerous interesting results and many open problems (for the current state of the field, see [Cornfeld *et al.*, 1982; Petersen, 1983; Mañé, 1987]). Yet the relevance of the enterprise for the foundations of statistical mechanics is often doubted. Thus [Earman and Rédei, 1996] argue that the enterprise is not relevant for explaining 'why phase averaging works' in equilibrium statistical mechanics; [Albert, 2000, p. 70] even calls the effort poured into rigorous proofs of ergodicity "nothing more nor less — from the standpoint of foundations of statistical mechanics — than a waste of time". (For further discussions, see: [Farquhar, 1964; Sklar, 1973; Friedman, 1976; Malament and Zabell, 1980; Leeds, 1989; van Lith, 2001a; Frigg, 2004; Berkovitz et al., 2006])

This judgment is usually based on the problems already indicated above; i.e. the difficulties of ascertaining that even the lowest property on the ergodic hierarchy actually obtains for interesting physical models in statistical mechanics, the empirical inaccessibility of infinite time averages, and the measure zero problem. Also, one often appeals to the Kolmogorov-Arnold-Moser (KAM) results[60] in order to temper the expectations that ergodicity could be a generic property of Hamiltonian systems. These difficulties are serious, but they do not, in my opinion, justify a definitive dismissal of ergodic theory.

Instead, it has been pointed out by [Khinchin, 1949; Malament and Zabell, 1980; Pitowsky, 2001] that further progress may be made by developing the theory in conditions in which (i) the equality of ensemble averages and time averages need not hold for *all* integrable functions, but for only a physically motivated subclass, (ii) imposing conditions that fix the rate of convergence in the infinite time limits in (99) and (103) and (iii) relaxing the conditions on what counts as an equilibrium state. Indeed important progress concerning (i) has been achieved in the 'theory of the thermodynamic limit', described in paragraph 6.3. It is clear that further alterations may be mathematically obstreperous; and that any results that might be obtained will not be as simple and general as those of the existing ergodic theory. But there is no reason why progress in these directions should be impossible. See e.g. [Vranas, 1998; van Lith, 2001b].

The measure zero problem, I would argue, is unsolvable within any "merely" measure-theoretic setting of the kind we have discussed above. The point is, that any measure theoretic discussion of dynamical systems that differ only on measure zero sets are, in measure-theoretical terms, isomorphic and usually identified. Measure theory has no way of distinguishing measure zero sets from the empty set. Any attempt to answer the measure zero problem should call upon other mathematical concepts. One can expect further light only by endowing the phase space with further physically relevant structure, e.g. a topology or a symplectic form (cf. [Butterfield, 2006; Belot, 2006]).

Furthermore, even if ergodic theory has little of relevance to offer to the explanation of 'why phase averaging works' in the case of equilibrium statistical mechanics, this does not mean it is a waste of time. Recall that the equality of phase and time averages was only one of several points on which the Ehrenfests argued that claims by Boltzmann could be substantiated by an appeal to the ergodic hypothesis. Another point was his (1877) claim that a system initially in a non-equilibrium macrostate should tend to evolve towards the equilibrium macrostate.

[60]Quite roughly, the KAM theorems show that some Hamiltonian systems for which trajectories are confined to an invariant set in phase space of small positive measure — and therefore *not* metrically transitive — , will continue to have that property when a sufficiently small perturbation is added to their Hamiltonian (for a more informative introduction, see [Tabor, 1989]). This conclusion spoilt the (once common) hope that non-metrically transitive systems were rare and idealized exceptions among Hamiltonian systems, and that they could always be turned into a metrically transitive system by acknowledging a tiny perturbation from their environment. As we have seen (p. 958), Boltzmann (1868) had already expressed this hope for the ergodic hypothesis.

It is ironic that some critics of ergodic theory dismiss the attempt to show in what sense and under which conditions the microstate does display a tendency to wander around the entire energy hypersurface as irrelevant, while relying on a rather verbal and pious hope that this will "typically" happen without any dynamical assumption to fall back on. Clearly, the ergodic hierarchy might still prove relevant here.

Still, it is undeniable that many concrete examples can be provided of systems that are not ergodic in any sense of the word and for which equilibrium statistical mechanics should still work. In a solid, say an ice cube, the molecules are tightly locked to their lattice site, and the phase point can access only a minute region of the energy hypersurface. Similarly, for a vapour/liquid mixture in a $\cap$-shaped vessel in a gravity field, molecules may spend an enormously long proportion of time confined to the liquid at the bottom of one leg of the vessel, even though the region corresponding to being located in the other leg is dynamically accessible. And still one would like to apply statistical mechanics to explain their thermal properties.

Summing up, even admitting that ergodic theory cannot provide the whole story in all desired cases does not mean it is irrelevant. I would argue that, on a qualitative and conceptual level, one of the most important achievements of ergodic theory is that it has made clear that strict determinism on the microscopic level is not incompatible with random behaviour on a macroscopic level, even in the strong sense of a Bernoulli system. This implies that the use of models with a stochastic evolution like urn drawings, that Boltzmann used in 1877, or the dog flea model of the Ehrenfests, (cf. §7.2), are not necessarily at odds with an underlying deterministic dynamics.

6.3 Khinchin's approach and the thermodynamic limit

In the 'hard core' version of ergodic theory, described in the previous two paragraphs, one focuses on abstract dynamical systems, i.e. the only assumptions used are about a measure space equipped with a dynamical evolution. It is not necessary that this dynamics arises from a Hamiltonian. Further, it is irrelevant in this approach whether the system has a large number of degrees of freedom. Indeed, the 'baker transformation', an example beloved by ergodic theorists because it provides a dynamical system that possesses *all* the properties distinguished in the ergodic hierarchy, uses the unit square as phase space, and thus has only two degrees of freedom. On the other hand, Hamiltonian systems with large numbers of degrees of freedom, may fail to pass even the lowest step of the ergodic hierarchy, i.e. metric transitivity.

This aspect of ergodic theory is often criticized, because the thermal behaviour of macroscopic systems that the foundations of statistical mechanics ought to explain, arguably appears only when their number of degrees of freedom is huge. As Khinchin puts it:

> All the results obtained by Birkhoff and his followers [...] pertain to

> the most general type of dynamic systems [...]. The authors of these studies have not been interested in the problem of the foundations of statistical mechanics which is our primary interest in this book. Their aim was to obtain the results in the most general form; in particular all these results pertain equally to the systems with only a few degrees of freedom as well as to the systems with a very large number of degrees of freedom.
>
> From our point of view we must deviate from this tendency. We would unnecessarily restrict ourselves by neglecting the special properties of the systems considered in statistical mechanics (first of all their fundamental property of having a very large number of degrees of freedom) [...]. Furthermore, we do not have any basis for demanding the possibility of substituting phase averages for the time averages of all functions; in fact the functions for which such substitution is desirable have many specific properties which make such a substitution apparent in these cases (Khinchin, 1949, p. 62).

Thus, partly in order to supplement, partly in competition to ergodic theory, Khinchin explored an approach to the ergodic problem that takes the large number of degrees of freedom as an essential ingredient, but only works for a specific class of functions, the so-called *sum functions.*

In particular, consider a Hamiltonian dynamical system $\langle\Gamma, \mathcal{A}, T, \mu\rangle$ of N point particles. That is, we assume: $x = (\vec{q}_1, \vec{p}_1; \ldots; \vec{q}_N, \vec{p}_N) \in \Gamma \subset \mathbb{R}^{6N}$. A function f on Γ is a sum function if

$$(114)\quad f(x) = \sum_{i=1}^{N} \phi_i(x_i)$$

where $x_i = (\vec{p}_i, \vec{q}_i)$ is the molecular state of particle i.[61] Under the further assumption that the Hamiltonian itself is a sum function, Khinchin proved:

> Khinchin's ergodic theorem: For all sum functions f there are positive constants κ_1, κ_2 such that, for all N:
>
> $$(115)\quad \mu\left(\left\{x \in \Gamma : \left|\frac{\overline{f(x)} - \langle f\rangle_\mu}{\langle f\rangle_\mu}\right| \geq \kappa_1 N^{-1/4}\right\}\right) \leq \kappa_2 N^{-1/4}$$

In words: as N becomes larger and larger, the measure of the set where $\bar{f}$ and $\langle f\rangle$ deviate more than a small amount goes to zero.

This theorem, then, provides an alternative strategy to address the ergodic problem: it says that time average and microcanonical phase average of sum functions will be roughly equal, at least in a very large subset of the energy hypersurface,

[61] Note that Khinchin does not demand that sum functions are symmetric under permutation of the particles.

provided that the number of particles is large enough. Of course, this 'rough equality' is much weaker than the strict equality 'almost everywhere' stated in the von Neumann-Birkhoff ergodic theorem. Moreover, it holds only for the sum functions (114). However, the assumption of metric transitivity is not needed here; nor is any of the more stringent properties of the ergodic hierarchy.

The advantages of this approach to the ergodic problem are clear: first, one avoids the problem that ergodic properties are hard to come by for physically interesting systems. Second, an important role is allotted to the large number of degrees of freedom, which, as noted above, seems a necessary, or at least welcome ingredient in any explanation of thermal behaviour,[62] and thirdly a physically motivated choice for special functions has been made.

However, there are also problems and drawbacks. First, with regard to the "infinite-times" problem (cf. p. 1010), Khinchin's approach fares no better or worse than the original ergodic approach. Second, since the rough equality does not hold "almost everywhere" but outside of a subset whose measure becomes small when N is large, the measure-zero problem of ergodic theory (p. 1010) is now replaced by a so-called "measure-epsilon problem": if we wish to conclude that in practice the time average and the phase average are (roughly) equal, we should argue that the set for which this does not hold, i.e. the set in the left-hand side of (115) is negligible. This problem is worse than the o measure-zero problem. For example, we cannot argue that ensembles whose density functions have support in such sets are excluded by an appeal to absolute continuity or translation continuity (cf. the discussion on p. 1003). Further, if we wish to apply the result to systems that are indeed not metrically transitive, there may be integrals of the equations of motion that lock the trajectory of the system into a tiny subset of Γ for all times, in which case such a set cannot be neglected for practical purposes (cf. [Farquhar, 1964]).

Khinchin argued that the majority of physically important phase functions that one encounters in statistical mechanics are sum functions (cf. [Khinchin, 1949, p. 63,97]). However, this view is clearly too narrow from a physical point of view. It means that all quantities that depend on correlations or interactions between the particles are excluded.

Finally there is the 'methodological paradox' [Khinchin, 1949, p. 41–43]. This refers to the fact that Khinchin had to assume that the Hamiltonian itself is also a sum function. Let me emphasize that this assumption is *not* made just for the purpose of letting the Hamiltonian be one of the functions to which the theorem applies; the assumption is crucial to the very derivation of the theorem. As Khinchin clearly notes, this is paradoxical because for an equilibrium state to arise at all, it is essential that the particles can interact (e.g. collide), while this possibility is denied when the Hamiltonian is a sum function.

In Khinchin's view, the assumption should therefore not be taken literally. In-

[62] The point can be debated, of course. Some authors argue that small systems can show thermal behaviour too, which statistical mechanics then ought explain. However, the very definition of thermal quantities (like temperature etc.) for such small systems is more controversial [Hill, 1987; Feshbach, 1987; Rugh, 2001; Gross and Votyakov, 2000].

stead, one should assume that there really are interaction terms in the Hamiltonian, but that they manifest themselves only at short distances between the particles, so that they can be neglected, except on a tiny part of phase space. Still, it remains a curious feature of his work that his theorem is intended to apply in situations that are inconsistent with the very assumptions needed to derive it (cf. [Morrison, 2000, p. 46-47]). As we shall see in the next paragraph, later work has removed this paradox, as well as many other shortcomings of Khinchin's approach.

The theory of the thermodynamic limit

The approach initiated by Khinchin has been taken further by van der Linde and Mazur (1963), and merged with independent work of van Hove, Yang and Lee, Fisher, Griffiths, Minlos, Ruelle, Lanford and others, to develop, in the late 60s and early 70s, into what is sometimes called the 'rigorous results' approach or the 'theory of the thermodynamic limit'. The most useful references are [Ruelle, 1969; Lanford, 1973; Martin-Löf, 1979]. The following is primarily based on [Lanford, 1973], which is the most accessible and also the most relevant for our purposes, since it explicitly addresses the ergodic problem, and on [van Lith, 2001b].

As in Khinchin's work, this approach aims to provide an explanatory programme for the thermal behaviour of macroscopic bodies in equilibrium by relying mostly on the following central points,

- One adopts the microcanonical measure on phase space.
- the observable quantities are phase functions F of a special kind (see below).
- The number of particles N is extremely large.

It is shown that, under some conditions, in the 'thermodynamic limit', to be specified below, the microcanonical probability distribution for F/N becomes concentrated within an narrow region around some fixed value. This result is similar to Khinchin's ergodic theorem. However, as we shall see, the present result is more powerful, while the assumptions needed are much weaker.

To start of, we assume a Hamiltonian, of the form

$$(116) \quad H(x) = \sum_{i}^{N} \frac{\vec{p}_i^2}{2m} + U(\vec{q}_1, \ldots, \vec{q}_N).$$

defined on the phase space Γ for N particles. For technical reasons, it is more convenient and simpler to work in the configuration space, and ignore the momenta. Consider a sequence of functions $F(\vec{q}_1, \ldots \vec{q}_n)$, $n = 1, 2, \ldots$ with an indefinite number of arguments, or, what amounts to the same thing, a single function F defined on

$$(117) \quad \cup_{n=1}^{\infty} (\mathbb{R}^3)^n.$$

Such a function is called an 'observable' if it possesses the following properties:

(a). Continuity: For each n, $F(\vec{q}_1, \dots \vec{q}_n)$ is a continuous function on $\mathbb{R}^{3n}$

(b). Symmetry: For each n, $F(\vec{q}_1, \dots \vec{q}_n)$ is invariant under permutation of its arguments.

(c). Translation invariance: For each n, and each $\vec{a} \in \mathbb{R}^3$, $F(\vec{q}_1 + \vec{a}, \dots \vec{q}_n + \vec{a}) = F(\vec{q}_1, \dots \vec{q}_n)$

(d). Normalization: $F(\vec{q}_1) = 0$

(e). Finite range: There exists a real number $R \in \mathbb{R}$ such that, for each n, the following holds: Suppose we divide the n particles into two clusters labeled by $i = 1, \dots m$, and $i' = 1, \dots m'$, where $m + m' = n$. If $|\vec{q}_i - \vec{q}_{i'}| > R$ for all i, i', then $F(\vec{q}_1, \dots \vec{q}_m; \vec{q}_1, \dots, \vec{q}_{m'}) = F(\vec{q}_1, \dots \vec{q}_m) + F(\vec{q}_1, \dots, \vec{q}_{m'})$.

For the most part, these conditions are natural and self-explanatory. Note that the symmetry condition (b) is very powerful. It may be compared to Boltzmann's (1877b) combinatorial approach in which it was argued that macrostates occupy an overwhelmingly large part of phases space due to their invariance under permutations of the particles (see §4.4). Note further that condition (e) implies that F reduces to a sum function if all particles are sufficiently far from each other. It also means that the observables characterized by Lanford may be expected to correspond to extensive quantities only. (Recall that a thermodynamical quantity is called extensive if it scales proportionally to the size of the system, and intensive if it remains independent of the system size.) In the present approach, intensive quantities (like temperature and pressure) are thus not represented as observables, but rather identified with appropriate derivatives of other quantities, after we have passed to the thermodynamical limit.

Further, it is assumed that the potential energy function U in (116) also satisfies the above conditions. In addition, the potential energy is assumed to be *stable*,[63] i.e.:

(f). Stability: There is a number $B \in \mathbb{R}$, such that, for all n and all $\vec{q}_1, \dots \vec{q}_n$:

(118) $U(\vec{q}_1, \dots \vec{q}_n) \geq -nB.$

This condition — which would be violated e.g. for Newtonian gravitational interaction — avoids that as n becomes large, the potential energy per particle goes to minus infinity, i.e., it avoids a collapse of the system.

For some results it is useful to impose an even stronger condition:

[63] Strictly speaking, condition (f) is not needed for the existence of the thermodynamic limit for the configurational microcanonical measure. It is needed, however, when these results are extended to phase space (or when using the canonical measure). Note also that the term "stability' here refers to an extensive lower bound of the Hamiltonian. This should be distinguished from thermodynamic concept of stability, which is expressed by the concavity of the entropy function (cf. p. 940).

(f′.) Superstability: The potential energy U is called superstable if, for every continuous function Φ of compact support in $\mathbb{R}^3$:

$$(119)\quad U(\vec{q}_1,\ldots \vec{q}_N) + \lambda \sum_{i\neq j} \Phi(\vec{q}_i - \vec{q}_j)$$

is stable for a sufficiently small choice of $\lambda > 0$. In other words, a stable potential is superstable if it remains stable when perturbed slightly by a continuous finite-range two-body interaction potential.

As in Khinchin's approach, the assumption (f) or (f′) is not just needed because one would like to count the potential energy among the class of observables; rather it is crucial to the proof of the existence of the thermodynamic limit. Of course, the assumption that the interaction potential is continuous and of finite range is still too restrictive to model realistic inter-molecular forces. As Lanford notes, one can weaken condition (e) to a condition of 'weakly tempered' potentials,[64], dropping off quickly with distance (cf. [Fisher, 1964, p. 386; Ruelle, 1969, p. 32] , although this complicates the technical details of the proofs. Again, it is clear, however, that some such condition on temperedness of the long range interactions is needed, if only to avoid another catastrophe, namely that the potential energy per particle goes to $+\infty$ as n increases, so that system might tend to explode. (As could happen, e.g. for a system of charges interacting by purely repulsive Coulomb forces.)

Now, with the assumptions in place, the idea is as follows. Choose a given potential U and an observable F obeying the above conditions. Pick two numbers u and ρ, that will respectively represent the (potential) energy per particle and the particle density (in the limit as N gets large), a bounded open region $\Lambda \subset \mathbb{R}^3$, and a large integer N, such that $\frac{N}{V(\Lambda)} \approx \rho$. (Here, $V(\Lambda)$ denotes the volume of Λ.) Further, choose a small number $\delta u > 0$, and construct the (thickened) energy hypersurface in configuration space, i.e. the shell:

$$(120)\quad \Omega_{\Lambda,N,u,\delta u} = \left\{ (\vec{q}_1,\ldots \vec{q}_N) \in \Lambda^N : \frac{U(\vec{q}_1,\ldots \vec{q}_N)}{N} \in (u-\delta u, u+\delta u) \right\}.$$

Let μ denote the Lebesgue measure on Λ^N; its (normalized) restriction to the above set may then be called the 'thickened configurational microcanonical measure'. Note that

$$(121)\quad \omega^{\rm cf}(E) := \int_{\Lambda^N} d\vec{q}_1 \cdots \vec{q}_N \, \delta(U(\vec{q}_1 \ldots \vec{q}_N) - E)$$

may be considered as the configurational analogue of the structure function (41). Thus

[64]If, for simplicity, the potential U is a sum of pair interactions $U = \sum_{i\neq j} \phi(\vec{q}_i - \vec{q}_j)$, it is weakly tempered iff there are real constants $R, D, \epsilon > 0$, such that $\phi(\vec{r}) \leq D\|\vec{r}\|^{3+\epsilon}$ when $\|\vec{r}\| \geq R$.

$$(122)\quad \mu(\Omega_{\Lambda,N,u,\delta u}) = \int_{\Lambda^N} d\vec{q}_1 \cdots \vec{q}_N \mathbb{1}_{(u-\delta u, u+\delta u)}(U/N) = \int_{N(u-\delta u}^{N(u+\delta u)} dE\, \omega^{\rm cf}(E),$$

so that $\frac{1}{2N\delta u}\mu(\Omega_{\Lambda,N,u,\delta u})$ provides a thickened or smoothened version of this configurational structure function. The reason for working with this thickened hypershell instead of the thin hypersurface is of course to circumvent singularities that may appear in the latter. In any case, we may anticipate that, when δu is small, this expression will represent the configurational part of the microcanonical entropy (84. A further factor $1/N!$ may be added to give this entropy a chance of becoming extensive.[65] (See also paragraph §5.2.

We are interested in the probability distribution of F/N with respect to this thickened microcanonical measure on configuration space. For this purpose, pick an arbitrary open interval J, and define

$$(123)\quad \mathcal{V}(\Lambda, N, u, \delta u, F, J) := \frac{1}{N!}\mu\left(\left\{(\vec{q}_1, \ldots, \vec{q}_N) \in \Omega_{u,\delta u} : \quad \frac{F(\vec{q}_1, \ldots, \vec{q}_N)}{N} \in J\right\}\right).$$

So,

$$(124)\quad \frac{1}{\mu(\Omega_{\Lambda,N,u,\delta u})}\mathcal{V}(\Lambda, N, u, \delta u, F, J) = \frac{\mathcal{V}(\Lambda, N, u, \delta u, F, J)}{\mathcal{V}(\Lambda, N, u, \delta u, F, \mathbb{R})}$$

gives the probability that F/N lies in the interval J with respect to the above microcanonical measure.

We wish to study the behaviour of this probability in the thermodynamic limit, i.e. as N becomes large, and $V(\Lambda)$ grows proportional to N, such that $N/V(\Lambda) = \rho$. This behaviour will depend on the precise details of the limiting procedure, in particular on the shape of Λ. Lanford chooses to take the limit in the sense of van Hove: A sequence of bounded open regions Λ in $\mathbb{R}^3$ is said to become infinitely large in the sense of Van Hove if, for all $r > 0$, the volume of the set of all points within a distance r from the boundary of Λ, divided by the volume of Λ, goes to zero as N goes to infinity. In other words, the volume of points close to the surface becomes negligible compared to the volume of the interior. This avoids that surface effects could play a role in the limiting behaviour — and eliminates the worry that interactions with the walls of the container should have been taken into account.

Now, the first major result is:

> (Existence of the thermodynamic limit.) As $N \longrightarrow \infty$, and Λ becomes infinitely large in the sense of Van Hove, in such a way that $N/V(\Lambda) = \rho$, then either of the following cases holds:
>
> (α). $\mathcal{V}(\Lambda, N, u, \delta u, F, J)$ goes to zero faster than exponentially in N, or:

[65] For example, if the system is an ideal gas, i.e. if $U(\vec{q}_1, \ldots, \vec{q}_N) \equiv 0$, one will have $\omega^{\rm cf}(E) = V^N = \left(\frac{N}{\rho}\right)^N$, so that $\ln \frac{1}{N!}\omega^{\rm cf}(E)$ scales proportionally to N, but $\ln \omega^{\rm cf}(E)$ does not.

(β). $\mathcal{V}(\Lambda, N, u, \delta u, F, J) \approx e^{Ns(\rho,u,\delta u,F,J)}$ where $s(\rho, F, J)$ does not depend on Λ or N, except through the ration $\frac{N}{V(\Lambda)} = \rho$.

In other words, this result asserts the existence of

$$(125)\quad s(\rho, u, \delta u, F, J) := \lim_{N \longrightarrow \infty} \frac{1}{N} \ln \mathcal{V}(\Lambda, N, u, \delta u, F, J)$$

where s is either finite or $-\infty$. (The possibility that $s = -\infty$ for *all* values of the arguments of s is further ignored.) This already gives some clue for how the probability (123) behaves as a function of J. If J_1 and J_2 are two open intervals, N is large, and we suppress the other variables for notational convenience, we expect:

$$(126)\quad \frac{\mu(\frac{F}{N} \in J_1)}{\mu(\frac{F}{N} \in J_2)} = \frac{\mathcal{V}(J_1)}{\mathcal{V}(J_2)} \approx e^{N(s(J_1)-s(J_2))}.$$

If $s(J_2) > s(J_1)$, this ratio goes to zero exponentially in N. Thus, for large systems, the probability $\mu(\frac{F}{N} \in J)$ will only be appreciable for those open intervals J for which $s(J)$ is large.

A stronger statement can be obtained as follows. Associated with the set function $s(J)$ one may define a point function s:

$$(127)\quad s(x) := \inf_{\substack{J \ni x \\ J \text{ open}}} s(J)$$

It can then be shown that, conversely, for all open J:

$$(128)\quad s(J) = \sup_{x \in J} s(x)$$

Moreover, — and this is the second major result — one can show:

(129) $s(x)$ is concave.

Further, $s(x)$ is finite on an open convex subset of its domain [Lanford, 1973, p. 26].

Now, it is evident that a concave function $s(x)$ may have three general shapes: It either achieves its maximum value: (i) never; (ii) exactly once, say in some point x_0; or (iii) on some interval. In case (i), F/N 'escapes to infinity' in the thermodynamic limit; this case can be excluded by imposing the superstability condition (f′). Case (ii) is, for our purpose, the most interesting one. In this case, we may consider intervals $J_2 = (x_0 - \epsilon, x_0 + \epsilon)$, for arbitrarily small $\epsilon > 0$ and J_1 any open interval that does not contain x_0; infer from (127,128) that $s(J_2) > s(J_1)$, and conclude from (126) that the relative probability for F/N to take a value in J_2 rather than J_1 goes to zero exponentially with the size of the system.

Thus we get the desired result: As N becomes larger and larger, the probability distribution of F/N approaches a delta function. Or in other words, the function F/N becomes roughly constant on an overwhelmingly large part of the configurational energy-hypershell:

$$(130) \quad \lim_{N\longrightarrow\infty} \mu\left(\left\{(\vec{q}_1,\ldots,\vec{q}_N)\in\Omega_{\Lambda,N,u,\delta u}: \quad |\frac{F(\vec{q}_1,\ldots,\vec{q}_N)}{N}-x_0|>\epsilon\right\}\right)=0$$

In case (iii), finally, one can only conclude that the probability distribution becomes concentrated on some interval, but that its behaviour inside this interval remains undetermined. One can show, if this interval is bounded, that this case is connected to phase transitions (but see [p. 12, 58 for caveats]). [66]

Remarks.

1. Phase transitions. First, it is obviously an immense merit of the theory of the thermodynamic limit that, in contrast to ergodic theory, it is, in principle, capable of explaining and predicting the occurrence of phase transitions from a model of the microscopic interaction, in further work often in conjunction with renormalization techniques. Indeed, this capability is its major claim to fame, quite apart from what it has to say about the ergodic problem. What is more, it is often argued that phase transitions are strictly impossible in any finite system, and thus absolutely *require* the thermodynamic limit [Styer, 2004; Kadanoff, 2000].

This argument raises the problem that our experience, including that of phase transitions in real physical bodies, always deals with finite systems. A theory that presents an account of phase transitions only in the thermodynamic limit, must then surely be regarded as an idealization. This conclusion will not come as a shock many physicists, since idealizations are ubiquitous in theoretical physics. Yet a curious point is that this particular idealization seems to be 'uncontrollable". See [Sklar, 2002] and [Liu, 1999; Callender, 2001; Batterman, 2005] for further discussion. I also note that an alternative approach has been proposed recently. In this view phase transitions are associated with topology changes in the microcanonical hypersurface $\{x : H(x) = E\}$ with varying E. The crucial distinction from the theory of the thermodynamic limit is, of course, is that such topology changes may occur in finite, — indeed even in small — systems (cf. [Gross, 1997; Gross and Votyakov, 2000; Casetti et al, 2003]). However this may be, I shall focus below on the virtues of the thermodynamic limit for the ergodic problem.

2. The ergodic problem revisited. When compared to ergodic theory or Khinchin's approach, the theory of the thermodynamic limit has much to speak in

[66]To see the connection (loosely), note that if one removes the condition $F/N \in J$ from the definition (123) — or equivalently, chooses $J = \mathbb{R}$ — , then s in (125) can be interpreted as the (thickened, configurational) microcanonical entropy per particle. Considered now as a function of the open interval $(u - \delta u, u + \delta u)$, s has the same properties as established for $s(J)$, since U itself belongs to the class of observables. Thus, here too, there exists a point function $s(u)$ analogous to (127), and this function is concave (Actually, if we restore one more variable in the notation, and write $s(\rho, u)$, the function is concave in both variables). In case (iii), therefore, this function is constant in u over some interval, say $[u'_0.u''_0]$. This means that there is then a range of thermodynamical states with the same temperature $T = (\frac{\partial s}{\partial u})^{-1}_\rho$, for a range of values of u and ρ, which is just what happens in the condensation phase transition in a van der Waals gas.

its favour. As in Khinchin's work, the problem of establishing metric transitivity for physically interesting systems does not arise, because the approach does not need to assume it. Further, as in Khinchin's work, the approach works only for special functions. But the class of functions singled out by the assumptions (a–f.) or (a–f′.) above is not restricted to (symmetric) sum functions, and allows for short-range interactions between the particles. Thus, unlike Khinchin, there is no methodological paradox (cf. p.1019).

Yet one might still question whether these assumptions are not too restrictive for physically interesting systems. On the one hand, it is clear that some conditions on temperedness and stability are needed to rule out catastrophic behaviour in the thermodynamic limit, such as implosion or explosion of the system. One the other hand, these assumptions are still too strong to model realistic thermal systems. The Coulomb interaction, which according to [Lieb and Lebowitz, 1973, p. 138] is "the true potential relevant for real matter", is neither tempered nor stable. A tour the force, performed by Lenard, Dyson, Lebowitz and Lieb, has been to extend the main results of the theory of the thermodynamical limit to systems interacting purely by Coulomb forces (if the net charge of the system is zero or small), both classically and quantum mechanically (for fermions) (see [Lieb, 1976, and literature cited therein]). This result, then, should cover most microscopic models of ordinary matter, as long as relativistic effects and magnetic forces can be ignored. But note that this extension is obtained by use of the canonical, rather than the microcanonical measure, and in view of the examples of non-equivalence of these ensembles (cf. p. 5.5) one might worry whether this result applies to ordinary matter in metastable states (like supersaturated vapours, and superheated or supercooled liquids).

Another remarkable point is that, unlike Khinchin's result (115), the result (130) does not refer to time averages at all. Instead, the *instantaneous* value of F/N is found to be almost constant for a large subset of the configurational energy hypersurface. Hence, there is also no problem with the infinite time limit (cf. p. 1010. Indeed, dynamics or time evolutions play no role whatsoever in the present results, and the contrast to the programme of ergodic theory is accordingly much more pronounced than in Khinchin's approach.

3. Problems left. What is left, in comparison to those two approaches to the ergodic problem, are two problems. First, there is still the question of how to motivate the choice for the configurational microcanonical measure (i.e. the normalized Lebesgue measure restricted to the energy hypershell). Lanford is explicit that the theory of the thermodynamic limit offers no help in this question:

> It is a much more profound problem to understand why events which are very improbable with respect to Lebesgue measure do not occur in nature. I, unfortunately, have nothing to say about this latter problem. [Lanford, 1973, p. 2].

For this purpose, one would thus have to fall back on other attempts at motivation (cf. p. 1003).

Secondly, there is the measure-epsilon problem (cf. p. 1019). The desired equality $F/N \approx x$ holds, according to (130), if N is large, outside of a set of small measure. Can we conclude that this set is negligible, or that its states do not occur in nature? In fact, the result (130) instantaneous values is so strong that one ought to be careful of not claiming too much. For example, it would be wrong to claim that for macroscopical systems (i.e. with $N \approx 10^{27}$), the set in the left-hand side of (130) does not occur in nature. Instead, it remains a brute fact of experience that macroscopic systems also occur in non-equilibrium states. In such states, observable quantities take instantaneous values that vary appreciably over time, and thus differ from their microcanonical average. Therefore, their microstate must then be located inside the set of tiny measure that one would like to neglect. Of course, one might argue differently if N is larger still, say $N = 10^{100}$ but this only illustrates the 'uncontrollability' of the idealization involved in this limit, i.e. one still lacks control over how large N must be to be sure that the thermodynamic limit is a reasonable substitute for a finite system.

Further points. Other points, having no counterpart in the approaches discussed previously, are the following. The approach hinges on a very delicately construed sequence of limits. We first have to take the thickened energy shell, then take N, Λ to infinity in the sense of van Hove, finally take δu to zero. But one may ask whether this is clearly and obviously the right thing to do, since there are alternative and non-equivalent limits (the sense of Fisher), the order of the limits clearly do not commute (the thickness of the energy hypershell is proportional to $N\delta u$), and other procedures like the 'continuum limit' [Compagner, 1989] have also been proposed.

Finally, in order to make full contact to classical statistical mechanics, on still has to lift restriction to configuration space, and work on phase space. [Lanford, 1973, p. 2] leaves this as a "straightforward exercise" to the reader. Let's see if we can fill in the details.

Suppose we start from a thickened microcanonical measure on phase space, with the same thickness $2N\delta u$, around a total energy value of $E_0 = Ne_0$. Its probability density is then given by

$$(131) \quad \rho_{Ne_0, N\delta u}(\vec{p}_1, \ldots \vec{p}_N; \vec{q}_1 \ldots \vec{q}_N) \quad = \quad \frac{1}{2N\delta u} \int_{E_0 - N\delta u}^{E_0 + N\delta u} \frac{1}{\omega(E)} \delta(H(x) - E) dE$$

For the Hamiltonian (116), the integral over the momenta can be performed (as was shown by Boltzmann [1868] (cf. Eqn (43). This yields a marginal density

$$(132) \quad \overline{\rho} Ne_0, N\delta u(\vec{q}_1, \ldots, \vec{q}_N) = \frac{1}{2N\delta u} \frac{2m\pi^{3N/2}}{\Gamma(\frac{3N}{2})} \int_{E_0 - N\delta u}^{E_0 + N\delta u} \frac{1}{\omega(E)} \left(2m(E - U(q))\right)^{(3N-2)/2} dE$$

This is not quite the normalized Lebesgue measure on configuration space employed by Lanford, but since the factor $(2m(E - U(q))^{(3N-2)/2}$ is a continuous function of U, — at least if $E_0 - N\delta u - U > 0$ — it is absolutely continuous with respect to the Lebesgue measure on the shell, and will converge to it in the limit $\delta u \longrightarrow 0$.

But in a full phase space setting, the physical quantities can also depend on the momenta, i.e., they will be functions $F(\vec{p}_1, \ldots \vec{p}_N; \vec{q}_1 \ldots \vec{q}_N)$ and, even if one assumes the same conditions (a–f) as before for their dependence on the second group of arguments, their probability distribution cannot always be determined from the configurational microcanonical measure. For example, let F_1 and F_2 be two observables on configuration space, for which F_1/N and F_2/N converge to different values in the thermodynamical limit, say x_1 and x_2, and let G be any symmetric function of the momenta that takes two different values each with probability 1/2. For example, take

$$(133)\quad G(\vec{p}_1, \ldots \vec{p}_N) = \begin{cases} 1 & \text{if } \sum_i \vec{p}_i \cdot \vec{n} \geq 0, \\ 0 & \text{elsewhere.} \end{cases},$$

for some fixed unit vector $\vec{n}$. Now consider the following function on phase space:

$$(134)\qquad A(\vec{p}_1, \ldots \vec{p}_N; \vec{q}_1 \ldots \vec{q}_N)) = \\ G(\vec{p}_1, \ldots \vec{p}_N)\, F_1(\vec{q}_1 \ldots \vec{q}_N) + G'(\vec{p}_1, \ldots \vec{p}_N)\, F_2(\vec{q}_1 \ldots \vec{q}_N),$$

where $G' = 1 - G$. If we first integrate over the momenta, we obtain $\tilde{A} = \frac{1}{2}(F_1 + F_2)$, which converges in the thermodynamical limit to $\frac{1}{2}(x_1 + x_2)$. However, it would be wrong to conclude that A is nearly equal to $\frac{1}{2}(x_1 + x_2)$ $(x_1 + x_2)/2$ in an overwhelmingly large part of phase space. Instead, it is nearly equal to x_1 on (roughly) half the available phase space and nearly equal to x_2 on the remaining half.

The extension of (130) to phase space functions will thus demand extra assumptions on the form of such functions; for example, that their dependence on the momenta comes only as some function of the kinetic energy, i.e.

$$(135)\quad A\vec{p}_1, \ldots \vec{p}_N; \vec{q}_1, \ldots \vec{q}_N) = \psi(\sum \frac{\vec{p}_i^2}{2m}) + F(\vec{q}_1, \ldots, \vec{q}_N)$$

for some continuous function ψ.

6.4 Lanford's approach to the Boltzmann equation

We now turn to consider some modern approaches to non-equilibrium statistical mechanics. Of these, the approach developed by Lanford and others (cf. [Lanford, 1975; Lanford, 1976; Lanford, 1981; Spohn, 1991; Cercignani *et al.*, 1994]) deserves special attention because it stays conceptually closer to Boltzmann's 1872 work on the Boltzmann equation and the H-theorem than any other modern approach to statistical physics. Also, the problem Lanford raised and tried to answer is one of no less importance than the famous reversibility and recurrence objections. Furthermore, the results obtained are the best efforts so far to show that a statistical reading of the Boltzmann equation or the H-theorem might hold for the hard spheres gas.

The question Lanford raised is that of the consistency of the Boltzmann equation and the underlying Hamiltonian dynamics. Indeed, if we consider the microstate of a mechanical system such as a dilute gas, it seems we can provide two competing accounts of its time evolution.

(1) On the one hand, given the mechanical microstate x_0 of a gas, we can form the distribution of state $f(\vec{r}, \vec{v})$, such that $f(\vec{r}, \vec{v})d^3\vec{v}d^3\vec{r}$ gives the relative number of molecules with a position between $\vec{r}$ and $\vec{r} + d^3\vec{r}$ and velocity between $\vec{v}$ and $\vec{v} + d^3\vec{v}$. Presumably, this distribution should be uniquely determined by the microstate x_0. Let us make this dependence explicit by adopting the notation $f^{[x_0]}$. This function, then, should ideally serve as an initial condition for the Boltzmann equation (48), and solving this equation — assuming, that is, that it, that it has a unique solution — would give us the shape of the distribution function at a later time, $f_t^{[x_0]}(\vec{r}, \vec{v})$.

(2) On the other hand, we can evolve the microstate x_0 for a time t with the help of the Hamiltonian equations. That will give us $x_t = T_t x_0$. This later state x_t will then also determine a distribution of state $f^{[x_t]}(\vec{r}, \vec{v})$.

It is a sensible question whether these two ways of obtaining a later distribution of state from an initial microstate are the same, i.e. whether the two time evolutions are consistent. In other words, the problem is whether the diagram below commutes:

$$\begin{array}{ccc} x_0 & \xrightarrow{\text{Hamilton}} & x_t \\ \downarrow & & \downarrow \\ f^{[x_0]} & \xrightarrow{\text{Boltzmann}} & f_t^{[x_0]} \stackrel{?}{=} f^{[x_t]} \end{array} \tag{136}$$

The first issue that has to be resolved here is the precise relation between a microstate and the distribution of state f. It is obvious that, in so far as this function represents the physical property of a gas system, it should be determined by the momentary microstate x. It is also clear, that in so far as it is assumed to be continuous and differentiable in time in order to obey the Boltzmann equation, this cannot be literally and exactly true.

So let us assume, as Boltzmann did, that the gas consists of N hard spheres, each of diameter d and mass m, contained in some fixed bounded spatial region Λ with volume $|\Lambda| = V$. Given a microstate x of the system one can form the 'exact' distribution of state:

$$F^{[x]}(\vec{r}, \vec{v}) := \frac{1}{N} \sum_i^N \delta^3(\vec{r} - \vec{q}_i)\delta^3(\vec{v} - \frac{\vec{p}_i}{m}). \tag{137}$$

This distribution is, of course, not a proper function, and being non-continuous and non-differentiable, clearly not a suitable object to plug into the Boltzmann equation. However, one may reasonably suppose that one ought to be able to express Boltzmann's ideas in a limit in which the number of particles, N, goes to infinity. However, this limit clearly must be executed with care.

On the one hand, one ought to keep the gas dilute, so that collisions involving three or more particles will be rare enough so that they can safely be ignored in comparison to two-particle collisions. On the other hand, the gas must not be so dilute that collisions are altogether too rare to contribute to a change of f. The appropriate limit to consider, as Lanford argues, is the so-called Boltzmann-Grad limit in which $N \longrightarrow \infty$, and:[67]

$$\frac{Nd^2}{V} = \text{constant} > 0. \tag{138}$$

Denote this limit as "$N \xrightarrow{\text{BG}} \infty$", where it is implicitly understood that $d \propto N^{-1/2}$. The hope is then that in this Boltzmann-Grad limit, the exact distribution $F^{[x^N]}$ will tend to a continuous function that can be taken as an appropriate initial condition for the Boltzmann equation. For this purpose, one has to introduce a relevant notion of convergence for distributions on the μ-space $\Lambda \times \mathbb{R}^3$. A reasonable choice is to say that an arbitrary sequence of distributions f_n (either proper density functions or in the distributional sense) converges to a distribution f, $f_n \longrightarrow f$, iff the following conditions hold:

For each rectangular parallelepiped $\Delta \subset \Lambda \times \mathbb{R}^3$:

$$\lim_{n \longrightarrow \infty} \int_\Delta f_N d^3\vec{r} d^3\vec{v} = \int_\Delta f d^3\vec{r} d^3\vec{v}, \tag{139}$$

and

$$\lim_{n \longrightarrow \infty} \int \vec{v}^2 f_n d^3\vec{r} d^3\vec{v} = \int \vec{v}^2 f d^3\vec{r} d^3\vec{v}, \tag{140}$$

where the second condition is meant to guarantee the convergence of the mean kinetic energy.

It is also convenient to introduce some distance function between (proper or improper) distributions that quantifies the sense in which one distribution is close to another in the above sense. That is to say, one might define some distance $d(f,g)$ between density functions on $\Lambda \times R^3$ such that

$$d(f_n, f) \longrightarrow 0 \Longrightarrow f_n \longrightarrow f. \tag{141}$$

There are many distance functions that could do this job, but I won't go into the question of how to pick out a particular one.

The hope is then, to repeat, that $F^{[x^N]} \longrightarrow f$ in the above sense when $N \xrightarrow{BG} \infty$, where f is sufficiently smooth to serve as an initial condition in the Boltzmann equation, and that with this definition, the Boltzmannian and Hamiltonian evolution become consistent in the sense that the diagram (136) commutes. But clearly this will still be a delicate matter. Indeed, increasing N means a transition from one mechanical system to another with more particles. But there is no obvious

[67]The condition can be explained by the hand-waving argument that Nd^2/V is proportional to the 'mean free path', i.e. a typical scale for the distance traveled by a particle between collisions, or also by noting that the collision integral in the Boltzmann equation is proportional to Nd^2/V, so that by keeping this combination constant, we keep the Boltzmann equation unchanged.

algorithm to construct the state x^{N+1} from x^N, and thus no way to enforce convergence on the level of individual states.

Still, one might entertain an optimistic guess, which, if true, would solve the consistency problem between the Boltzmann and the Hamiltonian evolution in an approximate fashion if N is very large.

> Optimistic Guess: If $F^{[x_0^N]}$ is near to f then $F^{[x_t^N]}$ is near to f_t for all $t > 0$, and where f_t is the solution of the Boltzmann equation with initial condition f.

As [Lanford, 1976] points out, the optimistic guess cannot be right. This is an immediate consequence of the reversibility objection: Indeed, suppose it were true for all $x \in \Gamma$, and $t > 0$. (Here, we momentarily drop the superscript N from x^N to relieve the notation.) Consider the phase point Rx obtained from x by reversing all momenta: $R(\vec{q}_1, \vec{p}_1; \ldots; \vec{q}_N, \vec{p}_N) = (\vec{q}_1, -\vec{p}_1; \ldots, ; \vec{q}_N, -\vec{p}_N)$. If $F^{[x]}(\vec{r}, \vec{v})$ is near to some distribution $f(\vec{r}, \vec{v})$, then $F^{[Rx]}(\vec{r}, \vec{v})$ is near to $f(\vec{r}, -\vec{v})$. But as x evolves to x_t, Rx_t evolves to $T_t Rx_t = RT_{-t}x_t = Rx$. Hence $F^{[T_t Rx_t]}(\vec{r}, \vec{v}) = F^{[Rx]}(\vec{r}, \vec{v})$ is near to $f(\vec{r}, -\vec{v})$. But the validity of the conjecture for Rx_t would require that $F^{[T_t Rx_t]}(\vec{r}, \vec{v})$ is near to $f_t(\vec{r}, -\vec{v})$ and these two distributions of state are definitely not near to each other, except in some trivial cases.

But even though the optimistic guess is false in general, one might hope that it is 'very likely' to be true, with some overwhelming probability, at least for some finite stretch of time. In order to make such a strategy more explicit, Lanford takes recourse to a probability measure on Γ, or more precisely a sequence of probability measures on the sequence of Γ_N's.

Apart from thus introducing a statistical element into what otherwise would have remained a purely kinetic theory account of the problem, there is a definite advantage to this procedure. As mentioned above, there is no obvious algorithm to construct a sequence of microstates in the Boltzmann-Grad limit. But for measures this is different. The microcanonical measure, for example is not just a measure for the energy hypersurface of one N-particles-system; it defines an algorithmic sequence of such measures for each N.

In the light of this discussion, we can now state Lanford's theorem as follows [Lanford, 1975; 1976]:

> Lanford's Theorem: Let $t \mapsto f_t$ be some solution of the Boltzmann equation, say for $t \in [0, a) \subset \mathbb{R}$. For each N, let Δ_N denote the set in the phase space Γ_N of N particles, on which $F^{[x^N]}$ is near to f_0 (the initial condition in the solution of the Boltzmann equation) in the sense that for some chosen distance function d and for tolerance $\epsilon > 0$:
>
> (142) $\Delta_N = \{x^N \in \Gamma_N \ : \ d(F^{[x^N]}, f_0) < \epsilon\}.$
>
> Further, for each N, conditionalize the microcanonical measure μ_N on Δ_N:

(143) $\mu_{\Delta,N}(\cdot) := \mu_N(\cdot|\Delta_N)$.

In other words, $\mu_{\Delta,N}$ is a sequence of measures on the various Γ_N that assign measure 1 to the set of microstates $x^N \in \Gamma_N$ that are close to f_0 in the sense that $d(F^{[x^N]}, f_0) < \epsilon$.
Then: $\exists\tau$, $0 < \tau < a$ such that for all t with $0 < t < \tau$:

(144) $\mu_{\Delta,N}(\{x^N \in \Gamma_N \ : \ d(F^{[x_t^N]}, f_t) < \epsilon\}) > 1 - \delta$

where $\delta \longrightarrow 0$ as both $\epsilon \longrightarrow 0$ and $N \overset{BG}{\longrightarrow} \infty$.

In other words: as judged from the microcanonical measure on Γ_N restricted to those states x^N that have their exact distribution of state close to a given initial function f_0, a very large proportion $(1 - \delta)$ evolve by the Hamiltonian dynamics in such a way that their later exact distribution of state $F^{[x_t^N]}$ remains close to the function f_t, as evolved from f_0 by the Boltzmann equation.

Remarks

Lanford's theorem shows that a statistical and approximate version of the Boltzmann equation can be derived from Hamiltonian mechanics and the choice of an initial condition in the Boltzmann-Grad limit. This is a remarkable achievement, that in a sense vindicates Boltzmann's intuitions. According to [Lanford, 1976, p. 14], the theorem says that the approximate validity of the Boltzmann equation, and hence the H-theorem, can be obtained from mechanics alone and a consideration of the initial conditions.

Still the result established has several remarkable features, all of which are already acknowledged by Lanford. First, there are some drawbacks that prevent the result from having practical impact for the project of justifying the validity of the Boltzmann equation in real-life physical applications. The density of the gas behaves like N/d^3, and in the Boltzmann-Grad limit this goes to zero. The result thus holds for extremely rarified gases. Moreover, the length of time for which the result holds, i.e. τ, depends on the constant in (138), which also provides a rough order of magnitude for the mean free path of the gas . It turns out that, by the same order of magnitude considerations, τ is roughly two fifths of the mean duration between collisions. This is a disappointingly short period: in air at room temperature and density, τ is in the order of microseconds. Thus, the theorem does not help to justify the usual applications of the Boltzmann equation to macroscopic phenomena which demand a much longer time-scale.

Yet note that the time scale is not trivially short. It would be a misunderstanding to say that the theorem establishes only the validity of the Boltzmann equation for times so short that the particles have had no chance of colliding: In two fifths of the mean duration between collisions, about 40 % of the particles have performed a collision.

Another issue is that in comparison with Boltzmann's own derivation no explicit mention seems to have been of the *Stoßzahlansatz.* In part this is merely apparent. In a more elaborate presentation (cf. Lanford 1975, 1976), the theorem is not presented in terms of the microcanonical measure, but an arbitrary sequence of measures ν_N on (the sequence of phase spaces) Γ_N. These measures are subject to various assumptions. One is that each ν_N should be absolutely continuous with respect to the microcanonical measure μ_N, i.e. ν_N should have a proper density function

(145) $d\nu_N(x) = n_N(x_1, \ldots x_N)dx_1 \cdots x_N$

where $x_i = (\vec{q_i}, \vec{p_i})$ denotes the canonical coordinates of particle i. Further, one defines, for each N and $m < N$, the reduced density functions by

(146) $n_N^{(m)}(x_1, \ldots x_m) := \dfrac{N!}{(N-m)!}\dfrac{1}{N^m}\displaystyle\int n_N(x_1, \ldots x_N)dx_{m+1} \cdots dx_N$

i.e. as (slightly renormalized) marginal probability distributions for the first m particles. The crucial assumption is now that

(147) $\displaystyle\lim_{N \xrightarrow{\text{BG}} \infty} n_N^{(m)}(x_1, \ldots x_m) = n^{(1)}(x_1) \cdots n^{(1)}(x_m)$

uniformly on compact subsets of $(\Lambda \times \mathbb{R}^3)^m$. This assumption (which can be shown to hold for the microcanonical measures) is easily recognized as a measure-theoretic analogy to the *Stoßzahlansatz.* It demands, in the Boltzmann-Grad limit, statistical independence of the molecular quantities for any pair or m-tuple of particles at time $t = 0$. As Lanford also makes clear, it is assumption (146) that would fail to hold if we run the construction of the reversibility objection; (i.e. if we follow the states x in Δ_N for some time t, $0t < \tau$, then reverse the momenta, and try to apply the theorem to the set $\Delta'_N = \{Rx_t \ : \ x \in \Delta_N\}$).

But another aspect is more positive. Namely: Lanford's theorem does not need to assume explicitly that the *Stoßzahlansatz* holds *repeatedly.* Indeed a remarkable achievement is that once the factorization condition (146) holds for time $t = 0$ it will also hold for $0 < t < \tau$, albeit in a weaker form (as convergence in measure, rather than uniform convergence). This is sometimes referred to as "propagation of chaos" [Cercignani *et al.*, 1994].

But the main conceptual problem concerning Lanford's theorem is where the apparent irreversibility or time-reversal non-invariance comes from. On this issue, various opinions have been expressed. [Lanford, 1975, p. 110] argues that irreversibility is the result of passing to the Boltzmann-Grad limit. Instead, [Lanford, 1976] argues that it is due to condition (146) plus the initial conditions (i.e.: $x_N \in \Delta_N$).

However, I would take a different position. The theorem equally holds for $-\tau < t < 0$, with the proviso that f_t is now a solution of the anti-Boltzmann equation. This means that the theorem is, in fact, invariant under time-reversal.

6.5 The BBGKY approach

The so-called BBGKY-hierarchy (named after Bogolyubov, Born, Green, Kirkwood and Yvon) is a unique amalgam of the description of Gibbs and the approach of Boltzmann. The goal of the approach is to describe the evolution of ensembles by means of reduced probability densities, and to see whether a Boltzmann-like equation can be obtained under suitable conditions — and thereby an approach to statistical equilibrium.

First, consider an arbitrary time-dependent probability density ρ_t. The evolution of ρ is determined via the Liouville-equation by the Hamiltonian:

$$(148)\quad \frac{\partial \rho_t}{\partial t} = \{H, \rho\}.$$

Central in the present approach is the observation that for relevant systems in statistical mechanics, this Hamiltonian will be symmetric under permutation of the particles. Indeed, the Hamiltonian for a system of N indistinguishable particles usually takes the form

$$(149)\quad H(\vec{q}_1, \vec{p}_1; \ldots; \vec{q}_N, \vec{p}_N) = \sum_{=1}^{N} \frac{\vec{p}_i^{\,2}}{2m} + \sum_{i}^{N} V(\vec{q}_i) + \sum_{i<j}^{N} \phi(\|q_i - \vec{q}_j\|)$$

where V is the potential representing the walls of the bounded spatial region Λ, say:

$$(150)\quad V(\vec{q}) = \begin{cases} 0 & \text{if } \vec{q} \in \Lambda \\ \infty & \text{elsewhere} \end{cases}$$

and ϕ the interaction potential between particle i and j. This is not only symmetric under permutation of the particle labels, but even has the more special property that it is a sum of functions that never depend on the coordinates of more than *two* particles. (cf. the discussion in §6.3.)

Let us again use the notation $x = (\vec{q}_1, \vec{p}_1; \ldots; \vec{q}_N, \vec{p}_N) = (x_1, \ldots, x_N)$; with $x_i = (\vec{q}_i, \vec{p}_i)$, and consider the sequence of reduced probability density functions, defined as the marginals of ρ:

$$(151)\quad \begin{aligned} \rho^{(1)}(x_1) &:= \int \rho_t(x)\, dx_2 \cdots x_N \\ &\vdots \\ \rho^{(m)}(x_1, \ldots, x_m) &= \int \rho_t(x)\, dx_{m+1} \cdots dx_N \end{aligned}$$

Here, $\rho^{(m)}$ gives the probability density that particles $1, \ldots, m$ are located at specified positions $\vec{q}_1, \ldots \vec{q}_m$ and moving with the momenta $\vec{p}_1, \ldots \vec{p}_m$, whereas all remaining particles occupy arbitrary positions and momenta.

Symmetry of the Hamiltonian need not imply symmetry of ρ. But one might argue that we may restrict ourselves to symmetric probability densities if *all* observable quantities are symmetric. In that case, it makes no observable difference when two or more particles are interchanged in the microstate and one may replace ρ by its average under all permutations without changing the expectation values of any observable quantity. However this may be, we now assume that ρ is, in fact, symmetric under permutations of the particle labels. In other words, from now on $\rho^{(m)}$ gives the probability density that any arbitrarily chosen set of m particles have the specified values for position and momentum.

The guiding idea is now that for relevant macroscopic quantities, we do not need the detailed form of the time evolution of ρ_t. Rather, it suffices to focus on no more than just a few marginals from the hierarchy (151). For example, suppose a physical quantity represented as a phase function A is a symmetric sum function on Γ:

$$(152)\quad A(x) = \sum_{i=1}^{N} A(x_i)$$

Then

$$(153)\quad \langle A\rangle = N\int A(x_1)\rho^{(1)}(x_1)\,dx_1$$

which is a considerable simplification. But this is not to say that we can compute the evolution of $\langle A\rangle$ in time so easily.

Consider in particular $\rho_t^{(1)}$ in (151). This is the *one-particle distribution function*: the probability that an arbitrary particle is in the one-particle state $(\vec{p},\vec{q})$. This distribution function is in some sense analogous to Boltzmann's f. But note: ρ_1 is a marginal probability distribution; it characterizes an ensemble, whereas f is (in this context) a stochastic variable, representing a property of a single gas:

$$(154)\quad f(\vec{r},\vec{v})) = \frac{1}{N}\sum_i \delta(\vec{q}_i-\vec{r})\delta(\vec{v}-\frac{\vec{p}_i}{m}).$$

How does $\rho_t^{(1)}$ evolve? From the Liouville-equation we get

$$(155)\qquad \frac{\partial\rho^{(1)}(x_1)}{\partial t} \;=\; \int\{H,\rho\}d^3\vec{p}_2\cdots\vec{p}_N d\vec{q}_2\cdots\vec{q}_N.$$

It is convenient here to regard the Poisson bracket as a differential operator on ρ, usually called the Liouville operator $\mathcal{L}$:

$$(156)\quad \mathcal{L}\rho := \sum_{i=1}^{N}\left(\frac{\partial H}{\partial\vec{q}_i}\cdot\frac{\partial}{\partial\vec{p}_i}-\frac{\partial H}{\partial\vec{p}_i}\cdot\frac{\partial}{\partial\vec{q}_i}\right)\rho.$$

For the Hamiltonian (149) this can be expanded as:

$$(157)\quad \mathcal{L} = \sum_{i}^{N} \mathcal{L}_i^{(1)} + \sum_{i<j}^{N} \mathcal{L}_{ij}^{(2)}$$

where

$$(158)\quad \mathcal{L}_i^{(1)} := \vec{p}_i \cdot \frac{\partial}{\partial \vec{q}_i}$$

and

$$(159)\quad \mathcal{L}_{ij}^{(2)} := \frac{\partial \phi_{ij}}{\partial \vec{q}_i} \cdot \left(\frac{\partial}{\partial \vec{p}_i} - \frac{\partial}{\partial \vec{p}_j} \right)$$

The evolution of $\rho^{(1)}$ is then given by:

$$(160)\qquad \frac{\partial \rho_t^{(1)}(x_1)}{\partial t} = \mathcal{L}_1^{(1)} \rho_t^{(1)}(x_1) + \int dx_2 \mathcal{L}_{12}^{(2)} \rho^{(2)}(x_1, x_2)$$

More generally, for higher-order reduced distribution functions $\rho^{(m)}$) $(m \geq 2)$, the evolution is governed by the equations:

$$\begin{aligned} \frac{\partial \rho_t^{(m)}(x_1, \ldots, x_m)}{\partial t} &= \sum_{i=1}^{m} \mathcal{L}_i^{(1)} \rho_t^{(m)}(x_1, \ldots x_m) + \sum_{i<j=1}^{m} \mathcal{L}_{ij}^{(2)} \rho_t^{(m)}(x_1, \ldots, x_m) \\ (161)\qquad &\quad + \sum_{i}^{m} \int dx_{m+1} \mathcal{L}_{i,m+1}^{(2)} \rho_t^{(m+1)}(x_1, \ldots x_{m+1}) \end{aligned}$$

The equations (160,161) form the *BBGKY hierarchy.* It is strictly equivalent to the Hamiltonian formalism for symmetric ρ and H, provided that H contains no terms that depend on three or more particles. As one might expect, solving these equations is just as hard as for the original Hamiltonian equations. In particular, the equations are not closed: in order to know how $\rho_t^{(1)}$ evolves, we need to know $\rho_t^{(2)}$. In order to know how $\rho_t^{(2)}$ evolves, we need to know $\rho_t^{(3)}$ etc.

The usual method to overcome this problem is to cut off the hierarchy, i.e. to assume that for some finite m, $\rho^{(m)}$ is a functional of $\rho^{(\ell)}$ with $\ell < m$. In particular, if we just consider the easiest case $(m = 2)$ and the easiest form of the functional, we can take $\rho^{(2)}$ to factorize in the distant past $(t \longrightarrow -\infty)$, giving:

$$(162)\quad \rho_t^{(2)}(x_1, x_2) = \rho_t^{(1)}(x_1)\rho_t^{(1)}(x_2); \quad \text{if } t \longrightarrow -\infty$$

i.e., requiring that the molecular states of any pair of particles are uncorrelated *before* their interaction. This is analogous to the *Stoßzahlansatz* (29), but now, of course, formulated in terms of the reduced distribution functions of an ensemble.

It can be shown that for the homogeneous case, i.e. when $\rho^{(2)}$ is uniform over the positions $\vec{q}_1$ and $\vec{q}_2$, i.e. when $\rho^{(2)}(x_1, x_2) = \rho^{(2)}(\vec{p}_1, \vec{p}_2)$ and when ϕ is a interaction potential of finite range, the evolution equation for $\rho^{(1)}$ becomes formally identical to the Boltzmann equation (48). That is to say, in (160) we may substitute $\mathcal{L}_i^{(1)} = 0$ and:

$$\begin{aligned} \frac{\partial \rho_t^{(1)}(\vec{p}_1)}{\partial t} &= \int \mathcal{L}_{12}^{(2)} \rho(\vec{p}_1, \vec{p}_2) d^3\vec{p}_2 \\ &= \frac{N}{m} \int bdbd\phi \int d\vec{p}_2 \|\vec{p}_2 - \vec{p}_1\| \\ &\quad \Big(\rho_t^{(1)}(\vec{p}_1{}')\rho_t^{(1)}(\vec{p}_2{}') - \rho_t^{(1)}(\vec{p}_1)\rho_t^{(1)}(\vec{p}_2)\Big) \end{aligned} \tag{163}$$

(See [Uhlenbeck and Ford, 1963, p. 131] for more details.)

Remarks

The BBGKY approach is thoroughly Gibbsian in its outlook, i.e. it takes a probability density over phase space as its basic conceptual tool. An additional ingredient, not used extensively by Gibbs, is its reliance on permutation symmetry. It gives an enormous extension of Gibbs' own work by providing a systematic hierarchy of evolution equations for reduced (or marginalized) density functions, which can then be subjected to the techniques of perturbation theory. An ensemble-based analogy of the Boltzmann equation comes out of this approach as a first-order approximation for dilute gases with collision times much smaller than the mean free time. The Boltzmann equation for inhomogeneous gases cannot be obtained so easily– as one might expect also on physical grounds that one will need extra assumptions to motivate its validity.

It is instructive to compare this approach to Lanford's. His analogy of the Boltzmann equation is obtained for a different kind of function, namely the one-particle distribution function $F^{[x]}$, i.e. the exact relative number of particles with molecular state $(\vec{r}, \vec{v})$, instead of $\rho^{(1)}$. Of course, there is a simple connection between the two. Noting that $F^{[x]}$ is a sum function (cf. equation (137), we see that

$$\langle F^{[x]} \rangle = \int \rho^{(1)}(\vec{p}_1, \vec{q}_1) f(\delta(\vec{r} - \vec{q}_1)\delta(\vec{v} - \frac{\vec{p}_1}{m}) dp_1 dq_1 = \rho^{(1)}(\vec{r}, \vec{v}). \tag{164}$$

In other words, the one-particle distribution function $\rho^{(1)}$ is the expected value of the exact distribution of state. It thus appears that where Lanford describes the probability of the evolution of the exact distribution of state, the BBGKY result (164) describes the evolution of the average of the exact distribution of state. Lanford's results are therefore much more informative.

One might be tempted here to argue that one can justify or motivate that that actual particle distribution might be taken equal to its ensemble average by arguments similar to those employed in ergodic theory. In particular, we have seen from Khinchin's work (cf. §6.3) that for large enough systems, the probability that a sum function such as $F^{[x]}$ deviates significantly from its expectation value is negligible. However, an important complication is that this reading of Khinchin's results holds for equilibrium, i.e. they apply with respect to the microcanonical distribution $\rho_{\rm mc}$, not to an arbitrary time-dependent density ρ_t envisaged here.

The time asymmetry of the resulting equation does not derive from the hierarchy of equations, but from the ensemble-based analogy of the *Stoßzahlansatz* (162). That is to say, in this approach time asymmetry is introduced via an initial condition on the ensemble, i.e. the absence of initial correlations. It can be shown, just like for the original Boltzmann equation, that when the alternative boundary condition is imposed that makes the momenta independent after collisions, (i.e. if (162) is imposed for $t \longrightarrow \infty$ instead) the anti-Boltzmann equation is obtained (see [Uhlenbeck and Ford, 1963, p. 127]).

7 STOCHASTIC DYNAMICS

7.1 Introduction

Over recent decades, some approaches to non-equilibrium statistical mechanics, that differ decidedly in their foundational and philosophical outlook, have nevertheless converged in developing a common unified mathematical framework. I will call this framework 'stochastic dynamics', since the main characteristic feature of the approach is that it characterizes the evolution of the state of a mechanical system as evolving under stochastic maps, rather than under a deterministic and time-reversal invariant Hamiltonian dynamics.[68]

The motivations for adopting this stochastic type of dynamics come from different backgrounds, and one can find authors using at least three different views.

1. "Coarse graining" (cf. [van Kampen, 1962; Penrose, 1970]): In this view one assumes that on the microscopic level the system can be characterized as a (Hamiltonian) dynamical system with deterministic time-reversal invariant dynamics. However, on the macroscopic level, one is only interested in the evolution of macroscopic states, i.e. in a partition (or coarse graining) of the microscopic phase space into discrete cells. The usual idea is that the form and size of these cells are chosen in accordance with the limits of our observational capabilities. A more detailed exposition of this view is given in §7.5.

On the macroscopic level, the evolution now need no longer be portrayed as deterministic. When only the macrostate of a system at an instant is given, it is in general not fixed what its later macrostate will be, even if the underlying microscopic evolution is deterministic. Instead, one can provide *transition probabilities*, that specify how probable the transition from any given initial macrostate to later macrostates is. Although it is impossible, without further assumptions, to say anything general about the evolution of the macroscopically characterized states, it is possible to describe the evolution of an ensemble or a probability distribution over these states, in terms of a *stochastic process*.

[68] Also, the name has been used in precisely this sense already by Sudarshan and coworkers, cf. [Sudarshan et al., 1961; Mehra and Sudarshan, 1972].

2. "Interventionism", "tracing" or "open systems" (cf. [Blatt, 1959; Davies, 1976; Lindblad, 1976; Lindblad, 1983; Ridderbos, 2002]): On this view, one assumes that the system to be described is not isolated but in interaction with the environment. It is assumed that the total system, consisting of the system of interest and the environment can be described as a (Hamiltonian) dynamical system with a time-reversal invariant and deterministic dynamics. If we represent the state of the system by $x \in \Gamma^{(\mathrm{s})}$ and that of the environment by $y \in \Gamma^{(\mathrm{e})}$, their joint evolution is given by a one-parameter group of evolution transformations, generated from the Hamiltonian equations of motion for the combined system: $U_t : (x, y) \mapsto U_t(x, y) \in \Gamma^{(s)} \times \Gamma^{(e)}$. The evolution of the state x in the course of time is obtained by projecting, for each t, to the coordinates of $U_t(x, y)$ in $\Gamma^{(\mathrm{s})}$; call the result of this projection x_t. Clearly, this reduced time evolution of the system will generally fail to be deterministic, e.g. the trajectory described by x_t in $\Gamma^{(\mathrm{s})}$ may intersect itself.

Again, we may hope that this indeterministic evolution can nevertheless, for an ensemble of the system and its environment, be characterized as a stochastic process, at least if some further reasonable assumptions are made.

3. A third viewpoint is to deny [Mackey, 1992; 2001], or to remain agnostic about [Streater, 1995], the existence of an underlying deterministic or time-reversal invariant dynamics, and simply regard the evolution of a system as described by a stochastic process as a new fundamental form of dynamics in its own right.

While authors in this approach thus differ in their motivation and in the interpretation they have of its subject field, there is, as we shall see, a remarkable unity in the mathematical formalism adopted for this form of non-equilibrium statistical mechanics. The hope, obviously, is to arrange this description of the evolution of mechanical systems in terms of a stochastic dynamics in such a way that the evolution will typically display 'irreversible behaviour': i.e. an 'approach to equilibrium', that a Boltzmann-like evolution equation holds, that there is a stochastic analogy of the H-theorem, etc. In short, one would like to recover the autonomy and irreversibility that thermal systems in non-equilibrium states typically display.

We will see that much of this can be achieved with relatively little effort once a crucial technical assumption is in place: that the stochastic process is in fact a homogeneous Markov process, or, equivalently, obeys a so-called master equation. Much harder are the questions of whether the central assumptions of this approach might still be compatible with an underlying deterministic time-reversal invariant dynamics, and in which sense the results of the approach embody time-asymmetry. In fact we shall see that conflicting intuitions on this last issue arise, depending on whether one takes a probabilistic or a dynamics point of view towards this formalism.

From a foundational point of view, stochastic dynamics promises a new approach to the explanation of irreversible behaviour that differs in interesting ways from the more orthodox Hamiltonian or dynamical systems approach. In that approach, any account of irreversible phenomena can only proceed by referring to special initial conditions or dynamical hypotheses. Moreover, it is well-known that an ensemble of such systems will conserve (fine-grained) Gibbs entropy so that the account cannot rely on this form of entropy for a derivation of the increase of entropy.

In stochastic dynamics, however, one may hope to find an account of irreversible behaviour that is not tied to special initial conditions, but one that is, so to say, built into the very stochastic-dynamical evolution. Further, since Liouville's theorem is not applicable, there is the prospect that one can obtain a genuine increase of Gibbs entropy from this type of dynamics.

As just mentioned, the central technical assumption in stochastic dynamics is that the processes described have the Markov property.[69] Indeed, general aspects of irreversible behaviour pour out almost effortlessly from the Markov property, or from the closely connected "master equation". Consequently, much of the attention in motivating stochastic dynamics has turned to the assumptions needed to obtain this Markov property, or slightly more strongly, to obtain a non-invertible Markov process [Mackey, 1992]. The best-known specimen of such an assumption is [van Kampen, 1962] "repeated randomness assumption". And similarly, critics of this type of approach [Sklar, 1993; Redhead, 1995; Callender, 1999] have also focused their objections on the question just how reasonable and general such assumptions are (cf. paragraph 7.5).

I believe both sides of the debate have badly missed the target. Many authors have uncritically assumed that the assumption of a (non-invertible) Markov process does indeed lead to non-time-reversal-invariant results. As a matter of fact, however, the Markov property (for invertible or non-invertible Markov processes) is time-reversal invariant. So, any argument to obtain that property need not presuppose time-asymmetry. In fact, I will argue that this discussion of irreversible behaviour as derived from the Markov property suffers from an illusion. It is due to the habit of studying the prediction of future states from a given initial state, rather than studying retrodictions towards an earlier state. As we shall see, for a proper description of irreversibility in stochastic dynamics one needs to focus on another issue, namely the difference between backward and forwards transition probabilities.

In the next subsections, I will first (§7.2) recall the standard definition of a homogeneous Markov process from the theory of stochastic processes. Subsection 7.3 casts these concepts in the language of dynamics, introduces the master equation, and discusses its analogy to the Boltzmann equation. In §7.4, we review some of the results that *prima facie* display irreversible behaviour for homogeneous Markov processes. In subsection 7.5 we turn to the physical motivations that have been given for the Markov property, and their problems, while §7.6 focuses on the question how seemingly irreversible results could have been obtained from a time-symmetric assumptions. Finally, §7.7 argues that a more promising discussion of these issues should start from a different definition of reversibility of stochastic processes.

---

[69] Some authors argue that the approach can and should be extended to include non-Markovian stochastic processes as well. Nevertheless I will focus here on Markov processes.

7.2 The definition of Markov processes

To start off, consider an example. One of the oldest discussions of a stochastic process in the physics literature is the so-called 'dog flea model' of P. and T. Ehrenfest (1907).

Consider N fleas, labeled from 1 to N, situated on either of two dogs. The number of fleas on dog 1 and 2 are denoted as n_1 and $n_2 = N - n_1$. Further, we suppose there is an urn with N lots carrying the numbers $1, \ldots N$ respectively. The urn is shaken, a lot is drawn (and replaced), and the flea with the corresponding label is ordered to jump to the other dog. This procedure is repeated every second.

It is not hard to see that this model embodies an 'approach equilibrium' in some sense: Suppose that initially all or almost all fleas are on dog 1. Then it is very probable that the first few drawings will move fleas from dog 1 to 2. But as soon as the number of fleas on dog 2 increases, the probability that some fleas will jump back to dog 1 increases too. The typical behaviour of, say, $|n_1 - n_2|$ as a function of time will be similar to Boltzmann's H-curve, with a tendency of $|n_1 - n_2|$ to decrease if it was initially large, and to remain close to the 'equilibrium' value $n_1 \approx n_2$ for most of the time. But note that in contrast to Boltzmann's H-curve in gas theory, the 'evolution' is here entirely stochastic, i.e. generated by a lottery, and that no underlying deterministic equations of motion are provided.

In general, a stochastic process is, mathematically speaking, nothing but a probability measure P on a measure space X, whose elements will here be denoted as ξ, on which there are infinitely many random variables Y_t, with $t \in \mathbb{R}$ (or sometimes $t \in \mathbb{Z}$). Physically speaking, we interpret t as time, and Y as the macroscopic variable(s) characterizing the macrostate — say the number of fleas on a dog, or the number of molecules with their molecular state in some cell of μ-space, etc. Further, ξ represents the total history of the system which determines the values of $Y_t(\xi)$. The collection Y_t may thus be considered as a single random variable Y evolving in the course of time.

At first sight, the name 'process' for a probability measure may seem somewhat unnatural. From a physical point of view it is the *realization*, in which the random variables Y_t attain the values $Y_t(\xi) = y_t$ that should be called a process. In the mathematical literature, however, it has become usual to denote the measure that determines the probability of all such realizations as a 'stochastic process'.

For convenience we assume here that the variables Y_t may attain only finitely many discrete values, say $y_t \in \mathcal{Y} = \{1, \ldots, m\}$. However, the theory can largely be set up in complete analogy for continuous variables.

The probability measure P provides, for $n = 1, 2, \ldots$, and instants $t_1, \ldots, t_n$ definite probabilities for the event that Y_t at these instants attains certain values $y_1, \ldots, y_n$:

$$\begin{gathered} P_{(1)}(y_1, t_1) \\ P_{(2)}(y_2, t_2; y_1, t_1) \\ \vdots \end{gathered}$$

$$P_{(n)}(y_n, t_n; \ldots; y_1, t_1) \tag{165}$$
$$\vdots$$

Here, $P_{(n)}(y_n, t_n; \ldots; y_1, t_1)$ stands for the joint probability that at times $t_1, \ldots, t_n$ the quantities Y_t attain the values $y_1, \ldots y_n$, with $y_i \in \mathcal{Y}$. It is an abbreviation for

$$P_{(n)}(y_n, t_n; \ldots; y_1, t_1) := P(\{\xi \in X \ : \ Y_{t_n}(\xi) = y_n \ \& \ \cdots \ \& \ Y_{t_1}(\xi) = y_1\}) \tag{166}$$

Obviously the probabilities (165) are normalized and non-negative, and each $P_{(n)}$ is a marginal of all higher-order probability distributions:

$$P_{(n)}(y_n, t_n; \ldots; y_1, t_1) = \sum_{y_{n+m}} \cdots \sum_{y_{n+1}} P_{(n+m)}(y_{n+m}, t_{n+m}; \ldots; y_1, t_1). \tag{167}$$

In fact, the probability measure P is uniquely determined by the hierarchy (165).[70]

Similarly, we may define conditional probabilities in the familiar manner, e.g.:

$$P_{(1|n-1)}(y_n, t_n | y_{n-1}, t_{n-1}; \ldots; y_1, t_1) := \frac{P_{(n)}(y_n, t_n; \ldots; y_1, t_1)}{P_{(n-1)}(y_{n-1}, t_{n-1}; \ldots; y_1, t_1)} \tag{168}$$

provides the probability that Y_{t_n} attains the value y_n, under the condition that $Y_{t_{n-1}}, \ldots, Y_{t_1}$ have the values $y_{n-1}, \ldots, y_1$.

In principle, the times appearing in the joint and conditional probability distributions (165,168) may be chosen in an arbitrary order. However, we adopt from now on the convention that they are ordered as $t_1 < \cdots < t_n$.

A special and important type of stochastic process is obtained by adding the assumption that such conditional probabilities depend only the condition at the last instant. That is to say: for all n and all choices of $y_1, \ldots y_n$ and $t_1 < \ldots < t_n$:

$$P_{(1|n)}(y_n, t_n | y_{n-1}, t_{n-1}; \ldots; y_n, t_n) = P_{(1|1)}(y_n, t_n | y_{n-1}, t_{n-1}) \tag{169}$$

This is the *Markov property* and such stochastic processes are called *Markov processes.*

The interpretation often given to this assumption, is that Markov processes have 'no memory'. To explain this slogan more precisely, consider the following situation. Suppose we are given a piece of the history of the quantity Y: at the instants $t_1, \ldots, t_{n-1}$ its values have been $y_1, \ldots, y_{n-1}$. On this information, we want to make a prediction of the value y_n of the variable Y at a later instant t_n. The Markov-property (169) says that this prediction would not have been better or worse if, instead of knowing this entire piece of prehistory, only the value y_{n-1} of Y at the last instant t_{n-1} had been given. Additional information about the past values is thus irrelevant for a prediction of the future value.

For a Markov process, the hierarchy of joint probability distributions (165) is subjected to stringent demands. In fact they are all completely determined by:

[70] At least, when we assume that the σ-algebra of measurable sets in X is the cylinder algebra generated by sets of the form in the right-hand side of (166).

(a) the specification of $P_{(1)}(y, 0)$ at one arbitrary chosen initial instant $t = 0$, and (b) the conditional probabilities $P_{(1|1)}(y_2, t_2|y_1, t_1)$ for all $t_2 > t_1$. Indeed,

$$P_{(1)}(y,t) = \sum_{y_0} P_{(1|1)}(y,t|y_0,0)P_{(1)}(y_0,0); \tag{170}$$

and for the joint probability distributions $P_{(n)}$ we find:

$$\begin{aligned} P_{(n)}(y_n,t_n;\ldots;y_1,t_1) &= P_{(1|1)}(y_n,t_n|y_{n-1},t_{n-1})\,P_{(1|1)}(y_{n-1},t_{n-1}|y_{n-2},t_{n-2}) \times \\ &\quad \times \cdots \times P_{(1|1)}(y_2,t_2|y_1,t_1)P_{(1)}(y_1,t_1). \end{aligned} \tag{171}$$

It follows from the Markov property that the conditional probabilities $P_{(1|1)}$ have the following property, known as the *Chapman-Kolmogorov* equation:

$$P_{(1|1)}(y_3,t_3|y_1,t_1) = \sum_{y_2} P_{(1|1)}(y_3,t_3|y_2,t_2)P_{(1|1)}(y_2,t_2|y_1,t_1) \quad \text{for } t_1 < t_2 < t_3. \tag{172}$$

So, for a Markov process, the hierarchy (165) is completely characterized by specifying $P_{(1)}$ at an initial instant and a system of conditional probabilities $P_{(1|1)}$ satisfying the Chapman-Kolmogorov equation. The study of Markov processes therefore focuses on these two ingredients.[71]

A following special assumption is *homogeneity*. A Markov process is called homogeneous if the conditional probabilities $P_{(1|1)}(y_2, t_2|y_1, t_1)$ do not depend on the two times t_1, t_2 separately but only on their mutual difference $t = t_2 - t_1$; i.e. if they are invariant under time translations. In this case we may write

$$P_{(1|1)}(y_2,t_2|y_1,t_1) = T_t(y_2,y_1) \tag{173}$$

such conditional probabilities are also called *transition* probabilities.

Is the definition of a Markov process time-symmetric? The choice in (169) of conditionalizing the probability distribution for Y_{t_n} on *earlier* values of Y_t is of course special. In principle, there is nothing in the formulas (165) or (168) that forces such an ordering. One might, just as well, ask for the probability of a value of Y_t in the past, under the condition that part of the *later* behaviour is given (or, indeed, conditionalize on the behaviour at both earlier and later instants.)

At first sight, the Markov property makes no demands about these latter cases. Therefore, one might easily get the impression that the definition is time-asymmetric. However, this is not the case. One can show that (169) is equivalent to:

$$P_{(1|n-1)}(y_1,t_1|y_2,t_2;\ldots;y_n,t_n) = P_{(1|1)}(y_1 t_1|y_2,t_2) \tag{174}$$

[71] Note, however, that although every Markov process is fully characterized by (i) an initial distribution $P_{(1)}(y, 0)$ and (ii) a set of transition probabilities $P_{(1|1)}$ obeying the Chapman-Kolmogorov equation and the equations (171), it is *not* the case that every stochastic process obeying (i) and (ii) is a Markov process. (See [van Kampen, 1981, p. 83] for a counterexample). Still, it is true that one can define a unique Markov process from these two ingredients by stipulating (171).

where the convention $t_1 < t_2 < \cdots < t_n$ is still in force. Thus, a Markov process does not only have 'no memory' but also 'no foresight'. Some authors (e.g. [Kelly, 1979]) adopt an (equivalent) definition of a Markov process that is explicitly time-symmetric: Suppose that the value y_i at an instant t_i somewhere in the middle of the sequence $t_1 < \cdots < t_n$ is given. The condition for a stochastic process to be Markov is then

$$\begin{aligned} &P_{(n|1)}(y_n, t_n; \ldots; y_1, t_1 | y_i, t_i) = \\ &\quad P_{(n-i|1)}(y_n, t_n; \ldots; y_{i+1}, t_{i+1} | y_i, t_i)\, P_{(i-1|1)}(y_{i-1}, t_{i-1}; y_1, t_1 | y_i, t_i) \end{aligned} \tag{175}$$

for all $n = 1, 2, \ldots$ and all $1 \leq i \leq n$. In another slogan: The future and past are independent if one conditionalizes on the present.

7.3 *Stochastic dynamics*

A homogeneous Markov process is for $t > 0$ completely determined by the specification of an initial probability distribution $P_{(1)}(y, 0)$ and the transition probabilities $T_t(y_2|y_1)$ defined by (173). The difference in notation (between P and T) also serves to ease a certain conceptual step. Namely, the idea is to regard T_t as a stochastic evolution operator. Thus, we can regard $T_t(y_2|y_1)$ as the elements of a matrix, representing a (linear) operator T that determines how an initial distribution $P_{(1)}(y, 0)$ will evolve into a distribution at later instants $t > 0$. (In the sequel I will adapt the notation and write $P_{(1)}(y, t)$ as $P_t(y)$.)

$$P_t(y) = (T_t P)(y) := \sum_{y'} T_t(y|y') P_0(y'). \tag{176}$$

The Chapman-Kolmogorov equation (172) may then be written compactly as

$$T_{t+t'} = T_t \circ T_{t'} \quad \text{for } t, t' \geq 0 \tag{177}$$

where $\circ$ stands for matrix multiplication, and we now also extend the notation to include the unit operator:

$$\mathbb{1}(y, y') = T_0(y, y') := \delta_{y,y'} \tag{178}$$

where δ denotes the Kronecker delta.

The formulation (177) can (almost) be interpreted as the group composition property of the evolution operators T. It may be instructive to note how much this is due to the Markov property. Indeed, for arbitrary conditional probabilities, say, if A_i, B_j and C_k denote three families of complete and mutually exclusive events (i.e. $\cup_i A_i = \cup_j B_j = \cup_k C_k = \mathcal{Y}$; $A_i \cap A_{i'} = B_j \cap B_{j'} = C_k \cap C_{k'} = \emptyset$ for $i \neq i', j \neq j'$ and $k \neq k'$), the rule of total probability gives :

$$P(A_i|C_k) = \sum_j P(A_i|B_j, C_k) P(B_j|C_k). \tag{179}$$

In general, this rule can *not* be regarded as ordinary matrix multiplication or a group composition! But the Markov property makes $P(A_i|B_j, C_k)$ in (179) reduce to $P(A_i|B_j)$, and then the summation in (179) coincides with familiar rule for matrix multiplication.

I wrote above: 'almost', because there is still a difference in comparison with the normal group property: in the Chapman-Kolmogorov-equation (177) all times must be positive. Thus, in general, for $t > 0$, T_t may not even be defined and so it does *not* hold that

$$(180) \quad T_{-t} \circ T_t = \mathbb{1}.$$

A family of operators $\{T_t, t \geq 0\}$ which is closed under a operation $\circ$ that obeys (177), and for which $T_0 = \mathbb{1}$ is called a *semigroup*. It differs from a group in the sense that its elements T_t need not be *invertible*, i.e., need not have an inverse. The lack of an inverse of T_t may be due to various reasons: either T_t does not possess an inverse, i.e. it is not a one-to-one mapping, or T_t does possess an inverse matrix T_t^{inv}, which however is itself non-stochastic (e.g. it may have negative matrix-elements). We will come back to the role of the inverse matrices in Sections 7.4 and 7.7.

The theory of Markov processes has a strong and natural connection with linear algebra. Sometimes, the theory is presented entirely from this perspective, and one starts with the introduction of a semigroup of *stochastic matrices*, that is to say, m by m matrices T with $T_{ij} \geq 0$ and $\sum_i T_{ij} = 1$. Or, more abstractly, one posits a class of states P, elements of a Banach space with a norm $\|P\|_1 = 1$, and a semigroup of stochastic maps T_t, $(t \geq 0)$, subject to the conditions that T_t is linear, positive, and preserves norm: $\|T_t P\|_1 = \|P\|_1$, (cf. [Streater, 1995]).

The evolution of a probability distribution P (now regarded as a vector or a state) is then particularly simple when t is discrete ($t \in \mathbb{N}$):

$$(181) \quad P_t = T^t P_0, \quad \text{where } T^t = \underbrace{T \circ \cdots \circ T}_{t \text{ times}}.$$

Homogeneous Markov processes in discrete time are also known as *Markov chains.*

Clearly, if we consider the family $\{T_t\}$ as a semigroup of stochastic evolution operators, or a stochastic form of dynamics, it becomes attractive to look upon $P_0(y)$ as a contingent initial state, chosen independently of the evolution operators T_t. Still, from the perspective of the probabilistic formalism with which we started, this might be an unexpected thought: both $P_{(1)}$ and $P_{(1|1)}$ are aspects of a single, given, probability measure P. The idea of regarding them as independent ingredients that may be specified separately doesn't then seem very natural. But, of course, there is no formal objection against the idea, since every combination of a system of transition probabilities T_t obeying the Chapman-Kolmogorov equation, and an arbitrary initial probability distribution $P_0(y) = P_{(1)}(y, 0)$ defines a unique homogeneous Markov process (cf. footnote 71). In fact, one sometimes even goes one step further and identifies a homogeneous Markov process completely with the specification of the transition probabilities, without regard of the initial state

$P_0(y)$; just like the dynamics of a deterministic system is usually presented without assuming any special initial state.

For Markov chains, the goal of specifying the evolution of $P_t(y)$ is now already completely solved in equation (181). In the case of continuous time, it is more usual to specify evolution by means of a differential equation. Such an equation may be obtained in a straightforward manner by considering a Taylor expansion of the transition probability for small times [van Kampen, 1981, p.101–103] — under an appropriate continuity assumption.

The result (with a slightly changed notation) is:

$$(182)\quad \frac{\partial P_t(y)}{\partial t} = \sum_{y'} \left(W(y|y')P_t(y') - W(y'|y)P_t(y)\right)$$

Here, the expression $W(y|y')$ is the transition probability from y' to y *per unit of time*. This differential equation, first obtained by Pauli in 1928, is called the *master equation*. (This name has become popular because an equation of this type covers a great variety of processes.)

The interpretation of the equation is suggestive: the change of the probability $P_t(y)$ is determined by making up a balance between gains and losses: the probability of value y increases in a time dt because of the transitions from y' to y, for all possible values of y'. This increase per unit of time is $\sum_{y'} W(y|y')P_t(y')$. But in same period dt there is also a decrease of $P_t(y)$ as a consequence of transitions from the value y to all other possible values y'. This provides the second term.

In this "balancing" aspect, the master equation resembles the Boltzmann equation (48), despite the totally different derivation, and the fact that $P_t(y)$ has quite another meaning than Boltzmann's $f_t(v)$. (The former is a probability distribution, the latter a distribution of particles.) Both are first-order differential equations in t. A crucial mathematical distinction from the Boltzmann equation is that the master equation is linear in P, and therefore much easier to solve.

Indeed, any solution of the master equation can formally be written as:

$$(183)\quad P_t = e^{tL}P_0,$$

where L represents the operator

$$(184)\quad L(y|y') := W(y|y') - \sum_{y''} W(y''|y')\delta_{y,y'}.$$

The general solution (183) is similar to the discrete time case (181), thus showing the equivalence of the master equation to the assumption of a homogeneous Markov process in continuous time.

A final remark(not needed for later paragraphs). The analogy with the Boltzmann equation can even be increased by considering a Markov process for particle pairs, i.e. by imagining a process where pairs of particles with initial states (i, j) make a transition to states (k, l) with certain transition probabilities (cf. [Alberti and Uhlmann, 1982, p. 30]) Let $W(i, j|k, l)$ denote the associated transition probability per unit of time. Then the master equation takes the form:

$$\text{(185)}\quad \frac{\partial P_t(i,j)}{\partial t} = \sum_{k,l} \left(W(i,j|k,l)P_t(k,l) - W(k,l|i,j)P_t(i,j) \right).$$

Assume now that the transitions $(i,j) \longrightarrow (k,l)$ and $(k,l) \longrightarrow (i,j)$ are equally probable, so that the transition probability per unit of time is symmetric: $W(i,j|k,l) = W(k,l|i,j)$, and, as an analogue to the *Stoßzahlansatz*, that $P(i,j)$ in the right-hand side may be replaced by the product of its marginals:

$$\text{(186)}\quad P(i,j) \longrightarrow \sum_j P(i,j) \cdot \sum_i P(i,j) = P'(i)P''(j)$$

Summing the above equation (185) over j, we finally obtain

$$\text{(187)}\quad \frac{\partial P'_t(i)}{\partial t} = \sum_j \frac{\partial P_t(i,j)}{\partial t} = \sum_{j,k,l} T(i,j|k,l)\big(P'_t(k)P''_t(l) - P'_t(i)P''_t(j)\big),$$

i.e., an even more striking analogue of the Boltzmann equation (48). But note that although (185) describes a Markov process, the last equation (187) does not: it is no longer linear in P, as a consequence of the substitution (186).

7.4 Approach to equilibrium and increase of entropy?

What can we say in general about the evolution of $P_t(y)$ for a homogeneous Markov process? An immediate result is this: the *relative entropy* is monotonically non-decreasing. That is to say, if we define

$$\text{(188)}\quad H(P,Q) := -\sum_{y\in\mathcal{Y}} P(y) \ln \frac{P(y)}{Q(y)}$$

as the relative entropy of a probability distribution P relative to Q, then one can show (see e.g. Moran 1961; Mackey 1991, p. 30):

$$\text{(189)}\quad H(P_t, Q_t) \geq H(P,Q)$$

where $P_t = T_t P$, $Q_t = T_t Q$, and T_t are elements of the semigroup (181) or (183).

One can also show that a non-zero relative entropy increase for at least some pair probability distributions P and Q, the stochastic matrix T_t must be non-invertible.

The relative entropy $H(P|Q)$ can, in some sense, be thought of as a measure of how much P and Q "resemble" each other.[72] Indeed, it takes its maximum value (i.e. 0) if and only if $P = Q$; it may become $-\infty$ if P and Q have disjoint support, (i.e. when $P(y)Q(y) = 0$ for all $y \in \mathcal{Y}$.) Thus, the result (189) says that if the stochastic process is non-invertible, pairs of distributions P_t and Q_t will generally become more and more alike as time goes by.

Hence it seems we have obtained a general weak aspect of "irreversible behaviour" in this framework. Of course, the above result does not yet imply that

[72] Of course, this is an asymmetric sense of "resemblance" because $H(P,Q) \neq H(Q,P)$.

the 'absolute' entropy $H(P) := -\sum_y P(y) \ln P(y)$ of a probability distribution is non-decreasing. But now assume that the process has a *stationary state*. In other words, there is a probability distribution $P^*(y)$ such that

(190) $T_t P^* = P^*.$

The intention is, obviously, to regard such a distribution as a candidate for the description of an equilibrium state. If there is such a stationary distribution P^*, we may apply the previous result and write:

(191) $H(P, P^*) \leq H(T_t P, T_t P^*) = H(P_t, P^*).$

In other words, as time goes by, the distribution $T_t P$ will then more and more resemble the stationary distribution than does P. If the stationary distribution is also uniform, i.e.:

(192) $P^*(y) = \dfrac{1}{m},$

then not only the relative but also the absolute entropy $H(P) := -\sum_y P(y) \ln P(y)$ increases, because

(193) $H(P, P^*) = H(P) - \ln m.$

In order to get a satisfactory description of an 'approach to equilibrium' the following questions remain:

(i) is there such a stationary distribution?

(ii) if so, is it unique?

(iii) does the monotonic behaviour of $H(P_t)$ imply that $\lim_{t \longrightarrow \infty} P_t = P^*$?

Harder questions, which we postpone to the next subsection 7.5, are:

(iv) how to motivate the assumptions needed in this approach or how to make judge their (in)compatibility with an underlying time deterministic dynamics; and

(v) how this behaviour is compatible with the time symmetry of Markov processes.

Ad (i). A stationary state as defined by (190), can be seen as an eigenvector of T_t with eigenvalue 1, or, in the light of (183), an eigenvector of L for the eigenvalue 0. Note that T or L are not necessarily Hermitian (or, rather, since we are dealing with real matrices, symmetric), so that the existence of eigenvectors is not guaranteed by the spectral theorem. Further, even if an eigenvector with the corresponding eigenvalue exists, it is not automatically suitable as a probability distribution because its components might not be positive.

Still, it turns out that, due to a theorem of Perron (1907) and Frobenius (1909), every stochastic matrix indeed has a eigenvector, with exclusively non-negative components, and eigenvalue 1 (see e.g. [Gantmacher, 1959; Van Harn and Holewijn, 1991]). But if the set $\mathcal{Y}$ is infinite or continuous this is not always true.

A well-known example of the latter case is the so-called Wiener process that is often used for the description of Brownian motion. It is characterized by the transition probability density:

(194) $T_t(y|y') = \frac{1}{\sqrt{2\pi t}} \exp \frac{(y-y')^2}{2t}, \quad y, y' \in \mathbb{R}.$

The evolution of an arbitrary initial probability density ρ_0 can be written as a convolution:

(195) $\rho_t(y) = \int T_t(y|y')\rho_0(y')dy';$

which becomes gradually lower, smoother and wider in the course of time, but does not approach any stationary probability density. Because this holds for every choice of ρ_0, there is no stationary distribution in this case.

However, it is not reasonable to see this as a serious defect. Indeed, in thermodynamics too one finds that a plume of gas emitted into free space will similarly diffuse, becoming ever more dilute without ever approaching an equilibrium state. Thermodynamic equilibrium is only approached for systems enclosed in a vessel of finite volume.

However, for continuous variables with a range that has finite measure, the existence of a stationary distribution is guaranteed under the condition that the probability density ρ_y is at all times bounded, i.e. $\exists M \in \mathbb{R}$ such that $\forall t \; \rho_t \leq M$; (see [Mackey, 1992, p. 36]).

Ad (ii). The question whether stationary solutions will be unique is somewhat harder to tackle. This problem exhibits an analogy to that of metric transitivity in the ergodic problem (cf. paragraph 6.1).

In general, it is very well possible that the range $\mathcal{Y}$ of Y can be partitioned in two disjoint regions, say A and B, with $\mathcal{Y} = A \cup B$, such that there are no transitions from A to B or vice versa (or that such transitions occur with probability zero). That is to say, the stochastic evolution T_t might have the property

(196) $T_t(Y \in A|Y \in B) = T_t(Y \in B|Y \in A) = 0$

In other words, its matrix may, (perhaps after a conventional relabeling of the outcomes) be written in the form:

(197) $\begin{pmatrix} T_A & 0 \\ 0 & T_B \end{pmatrix}.$

The matrix is then called (completely) *reducible.* In this case, stationary distributions will generally not be unique: If P^*_A is a stationary distribution with support in the region A, and P^*_B is a stationary distribution with support in B, then every convex combination

(198) $\alpha P^*_A(y) + (1-\alpha)P^*_B(y) \quad \text{with } 0 \leq \alpha \leq 1.$

will be stationary too. In order to obtain a unique stationary solution we will thus have to assume an analogue of metric transitivity. That is to say: we should demand that every partition of $\mathcal{Y}$ into disjoint sets A and B for which (196) holds is 'trivial' in the sense that $P(A) = 0$ or $P(B) = 0$.

So, one may ask, is the stationary distribution P^* unique if and only if the transition probabilities T_τ are not reducible? In the ergodic problem, as we saw in 6.1, the answer is positive (at least if P^* is assumed to be absolutely continuous with respect to the microcanonical measure). But not in the present case!

This has to do with the phenomenon of so-called 'transient states', which has no analogy in Hamiltonian dynamics. Let us look at an example to introduce this concept. Consider a stochastic matrix of the form:

$$(199) \quad \begin{pmatrix} T_A & T' \\ 0 & T_B \end{pmatrix}$$

where T' is a matrix with non-negative entries only. Then:

$$(200) \quad \begin{pmatrix} T_A & T' \\ 0 & T_B \end{pmatrix}\begin{pmatrix} P_A \\ 0 \end{pmatrix} = \begin{pmatrix} T_A P_A \\ 0 \end{pmatrix}, \qquad \begin{pmatrix} T_A & T' \\ 0 & T_B \end{pmatrix}\begin{pmatrix} 0 \\ P_B \end{pmatrix} = \begin{pmatrix} T' P_B \\ T_B P_B \end{pmatrix}$$

so that here transitions of the type $a \longrightarrow b$ have probability zero, but transitions of the type $b \longrightarrow a$ occur with positive probability. (Here, a, b stand for arbitrary elements of the subsets A and B.) It is clear that in such a case the region B will eventually be 'sucked empty'. That is to say: the total probability of being in region B (i.e. $\|T^t P_B\|$) will go exponentially to zero. The distributions with support in B are called 'transient' and the set A is called 'absorbing' or a 'trap'.

It is clear that these transient states will not play any role in the determination of the stationary distribution, and that for this purpose they might be simply ignored. Thus, in this example, the only stationary states are those with a support in A. And there will be more than one of them if T_A is reducible.

A matrix T that may be brought (by permutation of the rows and columns) in the form (199), with T_A reducible is called *incompletely reducible* [van Kampen, 1981, p. 108]. Further, a stochastic matrix is called *irreducible* if it is neither completely or incompletely reducible. An alternative (equivalent) criterion is that all states 'communicate' with each other, i.e. that for every pair of $i, j \in \mathcal{Y}$ there is some time t such that $P_t(j|i) > 0$.

The Perron-Frobenius theorem guarantees that as long as T irreducible, there is a unique stationary distribution. Furthermore, one can then prove an analogue of the ergodic theorem:[Petersen, 1983, p. 52]

> Ergodic theorem for Markov processes: If the transition probability T_t is irreducible, the time average of P_t converges to the unique stationary solution:
>
> $$(201) \quad \lim_{\tau\to\infty} \frac{1}{\tau}\int_0^\tau T_t P(y)dt = P^*(y).$$

Ad (iii). If there is a unique stationary distribution P^*, will $T_t P$ converge to P^*, for every choice of P? Again, the answer is not necessarily affirmative. (Even if (201) is valid!) For example, there are homogeneous and irreducible Markov chains for which P_t can be divided into two pieces: $P_t = Q_t + R_t$ with the following properties [Mackey, 1992, p. 71]:

1. Q_t is a term with $\|Q_t\| \longrightarrow 0$. This is a transient term.

2. The remainder R_t is periodic, i.e. after some finite time τ the evolution repeats itself: $R_{t+\tau} = R_\tau$.

These processes are called *asymptotically periodic*. They may very well occur in conjunction with a unique stationary distribution P^*, and show strict monotonic increase of entropy, but still not converge to P^*. In this case, the monotonic increase of relative entropy $H(P_t, P^*)$ is entirely due to the transient term. For the periodic piece R_t, the transition probabilities are permutation matrices, which, after τ repetitions, return to the unit operator.

Besides, if we arrange that P^* is uniform, we can say even more in this example: The various forms R_t that are attained during the cycle of permutations with period τ all have the same value for the relative entropy $H(R_t, P^*)$, but this entropy is strictly lower than $H(P^*, P^*) = 0$. In fact, P^* is the average of the R_t's, i.e.: $P^* = \frac{1}{\tau}\sum_{t=1}^{t=\tau} R_t$, in correspondence with (201).

Further technical assumptions can be introduced to block examples of this kind, and thus enforce a strict convergence towards the unique stationary distribution, e.g. by imposing a condition of 'exactness' [Mackey, 1992]. However, it would take us too far afield to discuss this in detail.

In conclusion, it seems that a weak aspect of "irreversible behaviour", i.e. the monotonic non-decrease of relative entropy is a general feature for all homogeneous Markov processes, (and indeed for all stochastic processes), and non-trivially so when the transition probabilities are non-invertible. Stronger versions of that behaviour, in the sense of affirmative answers to the questions (i), (ii) and (iii), can be obtained too, but at the price of additional technical assumptions.

7.5 Motivations for the Markov property and objections against them

Ad (iv). We now turn to the following problem: what is the motivation behind the assumption of the Markov property? The answer, of course, is going to depend on the interpretation of the formalism that one has in mind, and may be different in the 'coarse-graining' and the 'open systems' or interventionist approaches (cf. Section 7.1). I shall discuss the coarse-graining approach in the next paragraph below, and then consider the similar problem for the interventionist point of view.

Coarse-graining and the repeated randomness assumption

In the present point of view, one assumes that the system considered is really an isolated Hamiltonian system, but the Markov property is supposedly obtained from a partitioning of its phase space. But exactly how is that achieved?

One of the clearest and most outspoken presentations of this view is [van Kampen, 1962]. As in paragraph 5.4, we assume the existence of some privileged partition of the Hamiltonian phase space Γ — or of the energy hypersurface Γ_E — into disjoint cells: $\Gamma = \omega_1 \cup \cdots \cup \omega_m$. Consider an arbitrary ensemble with probability density ρ on this phase space. Its evolution can be represented by an operator

(202) $U_t^* \rho(x) := \rho(U_{-t}x)$,

where, — in order to avoid conflation of notation — we now use U_t to denote the Hamiltonian evolution operators, previously denoted as T_t, e.g. in (68) and throughout section 6. Let transition probabilities between the cells of this partition be defined as

(203) $$T_t(j|i) := P(x_t \in \omega_j | x \in \omega_i) = P(U_t x \in \omega_j | x \in \omega_i) = \frac{\displaystyle\int_{(U_{-t}\omega_j)\cap\omega_i} \rho(x)dx}{\int_{\omega_i} \rho(x)dx},$$

Obviously such transition probabilities will be homogeneous, due to the time-translation invariance of the Hamiltonian evolution U_t. Further, let $\hat{p}_0(i) := P(x \in \omega_i) = \int_{\omega_i} \rho(x)dx$, $i \in \mathcal{Y} = \{1, \ldots, m\}$, be an arbitrary initial coarse-grained probability distribution at time t=0.

Using the coarse-graining map defined by (92), one may also express the coarse-grained distribution at time t as

(204) $$\mathcal{CG}U_t^* \rho(x) = \sum_{ji} T_t(j|i)\hat{p}_0(i) \frac{1}{\mu(\omega_j)} \mathbb{1}_{\omega_j}(x)$$

where μ is the canonical measure on Γ, or the microcanonical measure on Γ_E. This expression indicates that, as long as we are only interested in the coarse grained history, it suffices to know the transition probabilities (203) and the initial coarse grained distributions.

But in order to taste the fruits advertised in the previous paragraphs, one needs to show that the transition probabilities define a Markov process, i.e., that they obey the Chapman-Kolmogorov equation (172),

(205) $T_{t'+t}(k|i) = T_{t'}(k|j)T_t(j|i)$; for all $t, t' > 0$.

Applying (204) for times t, t' and $t + t'$, it follows easily that the Chapman-Kolmogorov equation is equivalent to

(206) $\mathcal{CG}U_{t'+t}^* = \mathcal{CG}U_{t'}^* \mathcal{CG}U_t^*$, for all $t, t' > 0$.

In other words, the coarse-grained probability distribution at time $t + t'$ can be obtained by first applying the Hamiltonian dynamical evolution during a time t, then performing a coarse-graining operation, next applying the dynamical evolution during time t', and then coarse-graining again. In comparison to the relation $U^*_{t'+t} = U^*_{t'} U^*_t$, we see that the Chapman-Kolmogorov condition can be obtained by demanding that it is allowed to apply a coarse-graining, i.e. to reshuffle the phase points within each cell at any intermediate stage of the evolution. Of course, this coarse-graining halfway during the evolution erases all information about the past evolution apart from the label of the cell where the state is located at that time; and this ties in nicely with the view of the Markov property as having no memory (cf. the discussion on p. 1042).

What is more, the *repeated* application of the coarse-graining does lead to a monotonic non-decrease of the Gibbs entropy: If, for simplicity, we divide a time interval into m segments of duration τ, we have

$$(207)\quad \rho_{m\tau} = \underbrace{\mathcal{CG}U^*_\tau \, \mathcal{CG}U^*_\tau \cdots \mathcal{CG}U^*_\tau}_{m \text{ times}} \, \rho$$

and from (96):

$$(208)\quad \sigma[\rho_{m\tau}] \geq \sigma[\rho_{(m-1)\tau}] \geq \ldots \geq \sigma[\rho_\tau] \geq \sigma[\rho_0].$$

But since the choice of τ is arbitrary, we may conclude that $\sigma[\rho_t]$ is monotonically non-decreasing.

Thus, van Kampen argues, the ingredient to be added to the dynamical evolution is that, at any stage of the evolution, one should apply a coarse-graining of the distribution. It is important to note that it is not sufficient to do that just once at a single instant. At every stage of the evolution we need to coarse-grain the distribution again and again. Van Kampen [1962, p. 193] calls this the *repeated randomness* assumption.

What is the justification for this assumption? Van Kampen points out that it is "not unreasonable" (ibid., p. 182), because of the brute fact of its success in phenomenological physics. Thermodynamics and other phenomenological descriptions of macroscopic systems (the diffusion equation, transport equations, hydrodynamics, the Fokker-Planck equation, etc.) all characterize macroscopic systems with a very small number of variables. This means that their state descriptions are very coarse in comparison with the microscopic phase space. But their evolution equations are autonomous and deterministic: the change of the macroscopic variables is given in terms of the instantaneous values of those very same variables. The success of these equations shows, apparently, that the precise microscopic state does not add any relevant information beyond this coarse description. At the same time, van Kampen admits that the coarse-graining procedure is clearly not always successful. It is not difficult to construct a partition of a phase space into cells for which the Markov property fails completely.

Apparently, the choice of the cells must be "just right" [van Kampen, 1962, p. 183]. But there is as yet no clear prescription how this is to be done. Van

Kampen [1981, p. 80] argues that it is "the art of the physicist" to find the right choice, an art in which he or she succeeds in practice by a mixture of general principles and ingenuity, but where no general guidelines can be provided. The justification of the repeated randomness assumption is that it leads to the Markov property and from there onwards to the master equation, providing a successful autonomous, deterministic description of the evolution of the coarse-grained distribution.

It is worth noting that van Kampen thus qualifies the 'usual' point of view (cf. p. 977 above, and paragraph 5.4) on the choice of the cells; namely, that the cells are chosen in correspondence to our finite observation capabilities. Observability of the macroscopic variables is not sufficient for the success of the repeated randomness assumption. It is conceivable (and occurs in practice) that a particular partition in terms of observable quantities does not lead to a Markov process. In that case, the choice of observable variables is simply inadequate and has to be extended with other (unobservable) quantities until we (hopefully) obtain an exhaustive set, i.e. a set of variables for which the evolution can be described autonomously. An example is the spin-echo experiment: the (observable) total magnetization of the system does not provide a suitable coarse-grained description. For further discussion of this theme, see: [Blatt, 1959; Ridderbos and Redhead, 1998; Lavis, 2004; Balian, 2005].

Apart from the unsolved problem for which partition the repeated randomness assumption is to be applied, other objections have been raised against the repeated randomness assumption. Van Kampen actually gives us not much more than the advice to accept the repeated randomness assumption bravely, not to be distracted by its dubious status, and firmly keep our eyes on its success. For authors as [Sklar, 1993], who refers to the assumption as a "rerandomization posit", this puts the problem on its head. They request a justification of the assumption that would *explain* the success of the approach. (Indeed, even [van Kampen, 1981, p. 80] describes this success as a "miraculous fact"!). Such a request, of course, will not be satisfied by a justification that relies on its success. (But that does not mean, in my opinion, that it is an invalid form of justification.)

Another point that seems repugnant to many authors, is that the repeated coarse-graining operations appear to be added 'by hand', in deviation from the true dynamical evolution provided by U_t. The increase of entropy and the approach to equilibrium would thus apparently be a consequence of the fact that *we* shake up the probability distribution repeatedly in order to wash away all information about the past, while refusing a dynamical explanation for this procedure. [Redhead, 1995, p. 31] describes this procedure as "one of the most deceitful artifices I have ever come across in theoretical physics" (see also [Blatt, 1959] [Sklar, 1993] and [Callender, 1999] for similar objections).

One might ask whether the contrast between the repeated randomness assumption and the dynamical evolution need be so bleak as Van Kampen and his critics argue. After all, as we have seen in paragraph 6.2, there are dynamical systems so high in the ergodic hierarchy that they possess the Bernoulli property for some partition of phase space (cf. paragraph 6.2). Since the Markov property is weaker than the Bernoulli property, one may infer there are also dynamical systems whose coarse grained evolutions define a homogeneous Markov process.[73] Thus one might be tempted to argue that the Markov property, or the repeated randomness assumption proposed to motivate it, need not require a miraculous intervention from an external 'hand' that throws information away; a sufficiently complex deterministic dynamics on the microscopic phase space of the system might do the job all by itself. However, the properties distinguished in the ergodic hierarchy all rely on a given measure-preserving evolution. Thus, while some dynamical systems may have the Markov property, they only give rise to *stationary* Markov processes. Its measure-preserving dynamics still implies that the Gibbs entropy remains constant. Thus, the result (208) can only be obtained in the case when all inequality signs reduce to equalities. To obtain a non-trivial form of coarse-graining, we should indeed suspend the measure-preserving dynamics.

In conclusion,!!! although the choice of a privileged partition remains an unsolved problem, there need not be a conflict between the repeated randomness assumption and the deterministic character of the dynamics at the microscopic level. However, whether the assumption (206) might actually hold for Hamiltonian systems interesting for statistical mechanics is, as far as I know, still open.

Interventionism or 'open systems'

Another approach to stochastic dynamics is by reference to open systems. The idea is here that the system in continual interaction with the environment, and that this is responsible for the approach to equilibrium.

Indeed, it cannot be denied that in concrete systems isolation is an unrealistic idealization. The actual effect of interaction with the environment on the microscopic evolution can be enormous. A proverbial example, going back to [Borel, 1914], estimates the gravitational effect caused by displacing one gram of matter on Sirius by one centimeter on the microscopic evolution of an earthly cylinder of gas. Under normal conditions, the effect is so large, that, roughly and for a typical molecule in the gas, it may be decisive for whether or not this molecule will hit another given molecule after about 50 intermediary collisions. That is to say: microscopic dynamical evolutions corresponding to the displaced and the undisplaced matter on Sirius start to diverge considerably after a time of about 10^{-6} sec. In other words, the mechanical evolution of such a system is so extremely sensitive for disturbances of the initial state that even the most minute changes in the state of the environment can be responsible for large changes in the microscopic trajectory. But we cannot control the state of environment. Is it possible to regard irreversible behaviour as the result of such uncontrollable disturbances

[73]Strictly speaking this is true only for discrete dynamical systems. For continuous time, e.g. for Hamiltonian dynamics, the Markov property can only be obtained by adding a time smoothing procedure to the repeated randomness assumption [Emch, 1965],[Emch and Liu, 2001, pp. 484–486].

from outside?[74]

Let (x, y) be the state of a total system, where, as before, $x \in \Gamma^{(s)}$ represents the state of the object system and $y \in \Gamma^{(e)}$ that of the environment. We assume that the total system is governed by a Hamiltonian of the form

$$(209)\quad H_{\rm tot}(x,y) = H_{(s)} + H_{(e)} + \lambda H_{\rm int}(x,y),$$

so that the probability density of the ensemble of total systems evolves as

$$(210)\quad \rho_t(x,y) = U_t^* \rho_0(x,y) = \rho\left(U_{-t}(x,y)\right)$$

i.e., a time-symmetric, deterministic and measure-preserving evolution.

At each time, we may define marginal distributions for both system and environment:

$$(211)\quad \rho_t^{(s)}(x) = \int dy\, \rho_t(x,y),$$

$$(212)\quad \rho_t^{(e)}(x) = \int dx\, \rho_t(x,y).$$

We are, of course, mostly interested in the object system, i.e. in (211). Assume further that at time $t = 0$ the total density factorizes:

$$(213)\quad \rho_0(x,y) = \rho_0^{(s)}(x)\rho_0^{(e)}(y).$$

What can we say about the evolution of $\rho_t^{(s)}(x)$? Does it form a Markov process, and does it show increase of entropy?

An immediate result (see e.g. [Penrose and Percival, 1962]) is this:

$$(214)\quad \sigma[\rho_t^{(s)}] + \sigma[\rho_t^{(e)}] \geq \sigma[\rho_0^{(s)}] + \sigma[\rho_0^{(e)}],$$

where σ denotes the Gibbs fine-grained entropy (90). This result follows from the fact that $\sigma[\rho_t]$ is conserved and that the entropy of a joint probability distribution is always smaller than or equal to the sum of the entropies of their marginals; with equality if the joint distribution factorizes. This gives a form of entropy change for the total system, but it is not sufficient to conclude that the object system itself will evolve towards equilibrium, or even that its entropy will be monotonically increasing. (Notice that (214) holds for $t \leq 0$ too.)

Actually, this is obviously not to be expected. There are interactions with an environment that may lead the system away from equilibrium. We shall have to make additional assumptions about the situation. A more or less usual set of assumptions is:

(a). The environment is very large (or even infinite); i.e.: the dimension of $\Gamma^{(e)}$ is much larger than that of $\Gamma^{(s)}$, and $H_{(s)} \ll H_{(e)}$.

[74] Note that the term 'open system' is employed here for a system in (weak) interaction with its environment. This should be distinguished from the notion of 'open system' in other branches of physics where it denotes a system that can exchange particles with its environment.

(b). The coupling between the system and the environment is weak, i.e. λ is very small.

(c). The environment is initially in thermal equilibrium, e.g., $\rho^{(\mathrm{e})}(y)$ is canonical:

$$(215)\quad \rho_0^{(\mathrm{e})} = \frac{1}{Z(\beta)} e^{-\beta H^{(\mathrm{e})}}$$

(d). One considers time scales only that are long with respect to the relaxation times of the environment, but short with respect to the Poincaré recurrence time of the total system.

Even then, it is a major task to obtain a master equation for the evolution of the marginal state (211) of the system, or to show that its evolution is generated by a semigroup, which would guarantee that this forms a Markov process (under the proviso of footnote 71). Many specific models have been studied in detail (cf. [Spohn, 1980]). General theorems were obtained (although mostly in a quantum mechanical setting) by [Davies, 1974; Davies, 1976a; Lindblad, 1976; Gorini et al., 1976]. But there is a similarly to the earlier approach: it seems that, here too, an analogue of 'repeated randomness' must be introduced. [Mehra and Sudarshan, 1972; van Kampen, 1994; Maes and Netočný, 2003].

At the risk of oversimplifying and misrepresenting the results obtained in this analysis, I believe they can be summarized as showing that, in the so-called 'weak coupling' limit, or some similar limiting procedure, the time development of (211) can be modeled as

$$(216)\quad \rho_t^{(\mathrm{s})}(x) = T_t \rho^{(\mathrm{s})}(x) \quad t \geq 0,$$

where the operators T_t form a semigroup, while the environment remains in its steady equilibrium state:

$$(217)\quad \rho_t^{(\mathrm{e})}(y) = \rho_0^{(\mathrm{e})}(y) \quad t \geq 0.$$

The establishment of these results would also allow one to infer, from (214), the monotonic non-decrease of entropy of the system.

To assess these findings, it is convenient to define, for a fixed choice of $\rho_0^{(e)}$ the following linear map on probability distributions of the total system:

$$(218)\quad \mathcal{TR} : \rho(x,y) \mapsto \mathcal{TR}\rho(x,y) = \int \rho(x,y)dy \cdot \rho_0(y)$$

This map removes the correlation between the system and the environment, and projects the marginal distribution of the environment back to its original equilibrium form.

Now, it is not difficult to see that the Chapman-Kolmogorov equation (which is equivalent to the semigroup property) can be expressed as

$$(219)\quad \mathcal{TR}U^*_{t+t'} = \mathcal{TR}U^*_{t'}\mathcal{TR}U^*_t \quad \text{for all } t, t' \geq 0$$

which is analogous to (206).

There is thus a strong formal analogy between the coarse-graining and the open-systems approaches. Indeed, the variables of the environment play a role comparable to the internal coordinates of a cell in the coarse graining approach. The exact microscopic information about the past is here translated into the form of correlations with the environment. This information is now removed by assuming that at later times, effectively, the state may be replaced by a product of the form (213), neglecting the back-action of the system on the environment. The mappings $\mathcal{CG}$ and $\mathcal{TR}$ are both linear and idempotent mappings, that can be regarded as special cases of the projection operator techniques of Nakajima and Zwanzig, which allows for a more systematical and abstract elaboration, sometimes called *subdynamics*.

Some proponents of the open systems approach, (e.g. [Morrison, 1966; Redhead, 1995]), argue that in contrast to the coarse-graining approach, the present procedure is 'objective'. Presumably, this means that there is supposed to be a fact of the matter about whether the correlations are indeed 'exported to the environment'. However, the analogy between both approaches makes one suspect that any problem for the coarse-graining approach is translated into an analogous problem of the open systems approach. Indeed, the problem of finding a privileged partition that we discussed in the previous paragraph is mirrored here by the question where one should place the division between the 'system' and 'environment'. There is no doubt that it practical applications this choice is also arbitrary.

7.6 Can the Markov property explain irreversible behaviour?

Ad (v). Finally, I turn to what may well be the most controversial and surprising issue: is the Markov property, or the repeated randomness assumption offered to motivate it, responsible for the derivation of time-reversal non-invariant results?

We have seen that every non-invertible homogeneous Markov process displays "irreversible behaviour" in the sense that different initial probability distributions will tend to become more alike in the course of time. Under certain technical conditions, one can obtain stronger results, e.g. an approach to a unique equilibrium state, monotonic non-decrease of absolute entropy, etc. All these results seem to be clearly time-asymmetric. And yet we have also seen that the Markov property is explicitly time symmetric. How can these be reconciled?

To start off, it may be noted that it has often been affirmed that the Markov property is the key towards obtaining time-asymmetric results. For example, Penrose writes:

> " ... the behaviour of systems that are far from equilibrium is not symmetric under time reversal: for example: heat always flows from a hotter to a colder body, never from a colder to a hotter. If this behaviour could be derived from the symmetric laws of dynamics alone there would, indeed, be a paradox; we must therefore acknowledge the fact that some additional postulate, non-dynamical in character and

asymmetric under time reversal must be adjoined to the symmetric laws of dynamics before the theory can become rich enough to explain non-equilibrium behaviour. In the present theory, this additional postulate is the Markov postulate" (Penrose 1970, p. 41).

In the previous paragraph, we have already questioned the claim expressed here that the Markov property is "non-dynamical". But now we are interested in the question whether postulating the Markov property would be asymmetric under time-reversal. Many similar statements, e.g. that the repeated randomness assumption is "the additional element by which statistical mechanics has to be supplemented in order to obtain irreversible equations" [van Kampen, 1962, p. 182], or that the non-invertibility of a Markov process provides the origin of thermodynamic behaviour [Mackey, 1992] can be found in the works of advocates of this approach.

But how can this be, given that the Markov property is explicitly time-symmetric? In order to probe this problem, consider another question. How does a given probability distribution $P(y, 0)$ evolve for negative times? So, starting again from (170), let us now take $t \leq 0$. We still have:

(220) $$P(y,t) = \sum_{y'} P(y,t,|y',0)P(y',0).$$

These conditional probabilities $P(y, t, |y', 0)$ satisfy the 'time-reversed' Markov property (174), that says that extra specification of later values is irrelevant for the retrodiction of earlier values. As a consequence, we get for $t \leq t' \leq t'', 0$:

(221) $$P(y,t|y'',t'') \quad = \quad \sum_{y'} P(y,t|y',t')P(y',t'|y'',t'')$$

i.e., a time-reversed analogue of the Chapman-Kolmogorov equation.

We may thus also consider these conditional probabilities for negative times as backward evolution operators. If we could assume their invariance under time translation, i.e. that they depend only on the difference $\tau = t - t'$, we could write

(222) $$S_\tau(y|y') := P(y,t|y,t') \quad \text{with } \tau = t - t' \leq 0,$$

and obtain a second semigroup of operators S_τ, obeying

(223) $$S_{\tau+\tau'} = S_\tau \circ S_{\tau'} \quad \tau, \tau' \leq 0$$

that generate stochastic evolutions towards the past.

Further, these backward conditional probabilities are connected to the forward conditional probabilities by means of Bayes' theorem:

(224) $$P_{(1|1)}(y,t|y',t') = \frac{P_{(1|1)}(y',t'|y,t)P(y,t)}{P(y',t')};$$

and if the process, as before, is homogeneous this becomes

(225) $P_{(1|1)}(y,t|y',t') = \dfrac{T_{-\tau}(y'|y)P_t(y)}{P_{t'}(y')}$; $\tau = t - t' < 0.$

The matrix $P_{(1|1)}(y,t|y',t')$ always gives for $t < t'$ the correct 'inversion' of T_t. That is to say:

(226) $\sum_{y'} P(y,t|y',t')(T_{t'-t}P_t)(y') = P_t(y)$

Note firstly that (225) is *not* the matrix-inverse of T_t! Indeed, the right-hand side of (225) depends on P_t and $P_{t'}$ as well as T. Even if the matrix-inverse $T^{(\mathrm{inv})}$ does not exist, or is not a bona fide stochastic matrix, the evolution towards the past is governed by the Bayesian inversion, i.e. by the transition probabilities (225).

Note also that if the forward transition probabilities are homogeneous, this is not necessarily so for the backward transition probabilities. For example, if in (225) one translates both t and t' by δ, one finds

$$P(y,t+\delta|y',t'+\delta) \quad = \quad \frac{T_{-\tau}(y'|y)P(y,t+\delta)}{P(y',t'+\delta)}.$$

Here, the right-hand side generally still depends on δ. In the special case that the initial distribution is itself stationary, the backward transition probabilities are homogeneous whenever the forward ones are. If $P(y,t)$ is not stationary, we might still reach the same conclusion, as long as the non-stationarity is restricted to those elements y or y' of $\mathcal{Y}$ for which $T_t(y|y') = 0$ for all t. Otherwise, the two notions become logically independent.

This gives rise to an unexpected new problem. Usually, an assumption of homogeneity (or time translation invariance) is seen as philosophically innocuous, as compared to time reversal invariance. But here we see that assuming time translation invariance for a system of *forward* transition probabilities is not equivalent to assuming the same invariance for the *backward* transition probabilities. If one believes that one of the two is obvious, how will one go about explaining the failure of the other? And how would one explain the preference for which one of the two is obvious, without falling into the "double standards" accusation of the kind raised by [Price, 1996]?

But what about entropy increase? We have seen before that for every non-invertible Markov process the relative entropy of the distribution P with respect to the equilibrium distribution increases, and that the distribution evolves towards equilibrium. (Homogeneity of the process is not needed for this conclusion.) But the backward evolution operators form a Markov process too, for which exactly the same holds. This seems paradoxical. If $T_tP_0 = P_t$, we also have $P_t = S_{-t}P_0$. The entropy of P_t can hardly be both higher and lower than that of P_0! An example may clarify the resolution of this apparent problem: namely, the stationary solutions of S are not the same as the stationary solutions of T!

Example Consider a Markov chain with $\mathcal{Y} = \{1, 2\}$ and let

(227) $T = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}.$

Choose an initial distribution $P_0 = \begin{pmatrix} \alpha \\ 1-\alpha \end{pmatrix}$. After one step we already get

(228) $TP = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}$

which is also the (unique) stationary distribution P^*. The backward transition probabilities are given by Bayes' theorem, and one finds easily:

(229) $S = \begin{pmatrix} \alpha & \alpha \\ 1-\alpha & 1-\alpha \end{pmatrix}.$

The stationary distribution for this transition probability is

(230) $\tilde{P}^* = \begin{pmatrix} \alpha \\ 1-\alpha \end{pmatrix}.$

That is to say: for the forward evolution operator the transition

(231) $\begin{pmatrix} \alpha \\ 1-\alpha \end{pmatrix} \xrightarrow{T} \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}$

is one for which a non-stationary initial distribution evolves towards a stationary one. The relative entropy increases: $H(P_0, P^*) \leq H(TP, P^*)$. But for the backward evolution, similarly:

(232) $\begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix} \xrightarrow{S} \begin{pmatrix} \alpha \\ 1-\alpha \end{pmatrix}$

represents an evolution from a non-stationary initial distribution to the stationary distribution $\tilde{P}^*$ and, here too, relative entropy increases: $H(P_1, \tilde{P}^*) \leq H(P_0, \tilde{P}^*)$.

The illusion that non-invertible Markov processes possess a built-in time-asymmetry is (at least partly) due to the habit of regarding T_τ as a fixed evolution operator on an independently chosen distribution P_0. Such a view is of course very familiar in other problems in physics, where deterministic evolution operators generally *do* form a group and may be used, at our heart's desire, for positive and negative times.

Indeed, the fact that these operators in general have no inverse might seem to reflect the idea that Markov processes have no memory and 'loose information' along the way and that is the cause of the irreversible behaviour, embodied in the time-asymmetric master equation, increase of relative or absolute entropy or approach to equilibrium. But actually, every Markov process has apart from a system of forward, also a system of backward transition probabilities, that again forms a semigroup (when they are homogeneous). If we had considered *them* as given we would get all conclusions we obtained before, but now for negative times.

I conclude that irreversible behaviour is not built into the Markov property, or in the non-invertibility of the transition probabilities, (or in the repeated randomness assumption[75], or in the Master equation or in the semigroup property). Rather the appearance of irreversible behaviour is due to the choice to rely on the forward transition probabilities, and not the backward. A similar conclusion has been reached before [Edens, 2001] in the context of proposals of Prigogine and his coworkers. My main point here is that the same verdict also holds for more 'mainstream' approaches as coarse-graining or open systems.

7.7 Reversibility of stochastic processes

In order not to end this chapter on a destructive note, let me emphasize that I do not claim that the derivation of irreversible behaviour in stochastic dynamics is impossible. Instead, the claim is that motivations for desirable properties of the forward transition probabilities are not enough; one ought also show that these properties are lacking for the backward transitions.

In order to set up the problem of irreversibility in this approach to non-equilibrium statistical mechanics for a more systematic discussion, one first ought to provide a reasonable definition for what it means for a stochastic process to be (ir)reversible; a definition that would capture the intuitions behind its original background in Hamiltonian statistical mechanics.

One general definition that seems to be common (cf. [Kelly, 1979 p. 5]) is to call a stochastic process reversible iff, for all n and $t_1, \ldots, t_n$ and τ:

$$(233) \quad P_{(n)}(y_1, t_1; \ldots; y_n, t_n) = P_{(n)}(y_1, \tau - t_n; \ldots; y_n, \tau - t_n).$$

See [Grimmett and Stirzaker, 1982, p. 219] for a similar definition restricted to Markov processes) The immediate consequence of this definition is that a stochastic process can only be reversible if the single-time probability $P_{(1)}(y, t)$ is stationary, i.e. in statistical equilibrium. Indeed, this definition seems to make the whole problem of reconciling irreversible behaviour with reversibility disappear. As [Kelly, 1979, p. 19] notes in a discussion of the Ehrenfest model: "there is no conflict between reversibility and the phenomenon of increasing entropy — reversibility is a property of the model in equilibrium and increasing entropy is a property of the approach to equilibrium"

But clearly, this view trivializes the problem, and therefore it is not the appropriate definition for non-equilibrium statistical mechanics. Recall that the Ehrenfest dog flea model (§7.2) was originally proposed in an attempt of showing how a tendency of approaching equilibrium from a initial non-equilibrium distribution (e.g.

[75] In recent work, van Kampen acknowledges that the repeated randomness assumption by itself does not lead to irreversibility: "This repeated randomness assumption [...] breaks the time symmetry by explicitly postulating the randomization *at the beginning* of the time interval Δt. There is no logical justification for this assumption other than that it is the only thing one can do and that it works. If one assumes randomness at the end of each Δt coefficients for diffusion, viscosity, etc. appear with the wrong sign; if one assumes randomness at the midpoint no irreversibility appears" [van Kampen, 2002, p.475, original emphasis].

a probability distribution that gives probability 1 to the state that all fleas are located on the same dog) could be reconciled with a stochastic yet time-symmetric dynamics.

If one wants to separate considerations about initial conditions from dynamical considerations at all, one would like to provide a notion of (ir)reversibility that is associated with the stochastic dynamics alone, independent of the initial distribution is stationary.

It seems that an alternative definition which would fulfill this intuition is to say that a stochastic process is reversible if, for all y and y' and $t' > t$,

$$(234)\quad P_{(1|1)}(y,t|y',t') = P_{(1|1)}(y,t'|y',t).$$

In this case we cannot conclude that the process must be stationary, and indeed, the Ehrenfest model would be an example of a reversible stochastic process. I believe this definition captures the intuition that if at some time state y' obtains, the conditional probability of the state one time-step earlier being y is equal to that of the state one time-step later being y.

According to this proposal, the aim of finding the "origin" of irreversible behaviour or "time's arrow", etc. in stochastic dynamics must then lie in finding and motivating conditions under which the forward transition probabilities are different from the backwards transition probabilities, in the sense of a violation of (234). Otherwise, irreversible behaviour would essentially be a consequence of the assumptions about initial conditions, a result that would not be different in principle from conclusions obtainable from Hamiltonian dynamics.

BIBLIOGRAPHY

[Albert, 2000] D. Z. Albert. *Time and Chance* Cambridge, Mass.: Harvard University Press, 2000.

[Alberti and Uhlmann, 1982] P. M. Alberti amd A. Uhlmann. *Stochasticity and partial order. Doubly stochastic maps and unitary mixing.* Dordrecht: Reidel, 1982.

[Balescu, 1997] R. Balescu. *Statistical dynamics.* London: Imperial College Press, 1997.

[Balian, 2005] R. Balian. Information in statistical physics. *Studies In History and Philosophy of Modern Physics, 36*, 323–353, 2005.

[Batterman, 1990] R. W. Batterman. Irreversibility and statistical mechanics: A new approach? *Philosophy of Science, 57*, 395–419, 1990.

[Batterman, 1991] R. W. Batterman. Randomness and probability in dynamical theories: On the proposals of the Prigogine school. *Philosophy of Science, 58*, 241–263, 1991.

[Batterman, 2002] R. W. Batterman. *The devil in the details: asymptotic reasoning in explanation, reduction, and emergence* Oxford: Oxford University Press, 2002.

[Batterman, 2005] R. W. Batterman. Critical phenomena and breaking drops: Infinite idealizations in physics. *Studies in History and Philosophy of Modern Physics, 36*, 225–244, 2005.

[Becker, 1967] R. Becker. *Therory of heat* New York: Springer, 1967.

[Belot, 2006] G. Belot. This volume, chapter 2, 2006.

[Bennett, 2003] C. H. Bennett. Notes on Landauer's principle, reversible computation, and Maxwell's demon. *Studies In History and Philosophy of Modern Physics, 34*, 501–510, 2003.

[Berkovitz et al., 2006] Berkovitz, J., R. Frigg and F. Kronz. The ergodic hierarchy and Hamiltonian chaos. *Studies in History and Philosophy of Modern Physics*, 2006.

[Bertrand, 1889] J. Bertrand. *Calcul des Probabilités*. Paris: Gauthier-Villars. Reissued New York: Chelsea, 1889. no date.

[Bierhalter, 1994] G. Bierhalter. Von L. Boltzmann bis J.J. Thomson: die Versuche einer mechanischen Grundlegung der Thermodynamik (1866-1890). *Archive for History of Exact Sciences, 44*, 25–75, 1994.

[Birkhoff, 1931] G. D. Birkhoff. Proof of the ergodic theorem. *Proceedings of the National Academy of Sciences of the United States of America, 17*, 656–660, 1931.

[Bishop, 2004] R. C. Bishop. Nonequilibrium statistical mechanics Brussels-Austin style *Studies In History and Philosophy of Modern Physics, 35*, 1–30, 2004.

[Blatt, 1959] J. M. Blatt. An alternative approach to the ergodic problem. *Progress in Theoretical Physics, 22*, 745–756, 1959.

[Boltzmann, 1866] L. Boltzmann. Über die Mechanische Bedeutung des Zweiten Hauptsatzes der Wärmetheorie. *Wiener Berichte, 53*, 195–220 1866; in [Boltzmann, 1909] Vol. I, paper 2,.

[Boltzmann, 1868a] L. Boltzmann. Studien über das Gleichgewicht der lebendigen Kraft zwischen bewegten materiellen Punkten. *Wiener Berichte, 58*, 517–560, 1868; in [Boltzmann, 1909] Vol. I, paper 5.

[Boltzmann, 1868b] L. Boltzmann. Lösung eines mechanischen Problems. *Wiener Berichte, 58*, 1035–1044, 1868. In [Boltzmann, 1909] Vol. I, paper 6.

[Boltzmann, 1871a] L. Boltzmann. Über das Wärmegleichgewicht zwischen mehratomigen Gasmolekülen. *Wiener Berichte, 63*, 397–418, 1871. In [Boltzmann, 1909] Vol. I, paper 18.

[Boltzmann, 1871b] L. Boltzmann. Einige allgemeine Sätze über Wärmegleichgewicht. *Wiener Berichte, 63*, 679–711, 1871. In [Boltzmann, 1909] Vol. I, paper 19.

[Boltzmann, 1871c] L. Boltzmann. Analytischer Beweis des zweiten Haubtsatzes der mechanischen Wärmetheorie aus den Sätzen über das Gleichgewicht der lebendigen Kraft. *Wiener Berichte, 63*, 712–732, 1871. In [Boltzmann, 1909] Vol. I, paper 20.

[Boltzmann, 1872] L. Boltzmann. Weitere Studien über das Wärmegleichgewicht unter Gasmolekülen. *Wiener Berichte, 66*, 275–370, 1872. In [Boltzmann, 1909] Vol. I, paper 23.

[Boltzmann, 1877a] L. Boltzmann. . Bermerkungen über einige Probleme der mechanische Wärmetheorie *Wiener Berichte, 75*, 62–100, 1877. In [Boltzmann, 1909] Vol. II, paper 39.

[Boltzmann, 1877b] L. Boltzmann. Über die Beziehung zwisschen dem zweiten Haubtsatze der mechanischen Wärmetheorie und der Wahrscheinlichkeitsrechnung resp. dem Sätzen über das Wärmegleichgewicht *Wiener Berichte, 76*, 373–435, 1877. In [Boltzmann, 1909] Vol. II, paper 42.

[Boltzmann, 1878] L. Boltzmann. Weitere Bemerkungen über einige Probleme der mechanischen Wärmetheorie. *Wiener Berichte, 78*, 7–46, 1878. In [Boltzmann, 1909] Vol. II paper 44.

[Boltzmann, 1881a] L. Boltzmann. Über einige das Wärmegleichgewicht betreffende Sätze. *Wiener Berichte, 84* 136–145, 1881. In [Boltzmann, 1909] Vol. II paper 62.

[Boltzmann, 1881b] L. Boltzmann. Referat über die Abhandlung von J.C. Maxwell: "Über Boltzmann's Theorem betreffend die mittlere verteilung der lebendige Kraft in einem System materieller Punkte". *Wiedemann's Annalen Beiblätter, 5*, 403–417, 1881. In [Boltzmann, 1909] Vol. II paper 63.

[Boltzmann, 1884] L. Boltzmann. Über die Eigenschaften Monocyklischer und andere damit verwandter Systeme. *Crelle's Journal für die reine und angewandte Mathematik, 98*, 68–94 1884 and 1885. In [Boltzmann, 1909] Vol III, paper 73.

[Boltzmann, 1887a] L. Boltzmann. Ueber die mechanischen Analogien des zweiten Hauptsatzes der Thermodynamik. *Journal für die reine und angewandte Mathematik, 100*, 201–212, 1887. Also in [Boltzmann, 1909], Vol. III, paper 78.

[Boltzmann, 1887b] L. Boltzmann. . Neuer Beweis zweier Sätze über das Wärmegleichgewicht unter mehratomigen Gasmolekülen. *Wiener Berichte* **95**, 153–164, 1887 in [Boltzmann, 1909] Vol. III, paper 83.

[Boltzmann, 1887c] L. Boltzmann. Über einige Fragen der Kinetische Gastheorie. *Wiener Berichte* **96**, 891–918, 1887. In [Boltzmann, 1909] Vol. III, paper 86.

[Boltzmann, 1892] L. Boltzmann. III. Teil der Studien über Gleichgewicht der lebendigen Kraft *Münch. Ber. 22*, 329–358, 1892. In [Boltzmann, 1909] Vol. III, paper 97.

[Boltzmann, 1894a] L. Boltzmann and C.H. Bryan. Über die mechanische Analogie des Wärmegleichgewichtes zweier sich berührende Körper. *Wiener Berichte, 103*, 1125–1134, 1894. In [Boltzmann, 1909] Vol. III, paper 107.

[Boltzmann, 1894b] L. Boltzmann. . On the application of the determinantal relation to the kinetic theory of gases. Appendix C to an article by G.H. Bryan on thermodynamics. *Reports of the British Association for the Advancement of Science*, pp. 102–106, 1894. In [Boltzmann, 1909] Vol. III, paper 108.

[Boltzmann, 1895] L. Boltzmann. On certain questions in the theory of gases. *Nature* **51**, 413–415, 1895. In [Boltzmann, 1909], Vol. III, paper 112.

[Boltzmann, 1895b] L. Boltzmann. On the minimum theorem in the theory of gases. *Nature* **52**, 221, 1895. Also in [Boltzmann, 1909] Vol. III, paper 114.

[Boltzmann, 1896] L. Boltzmann. *Vorlesungen über Gastheorie* Vol I. Leipzig, J.A. Barth, 1896. Translated, together with [Boltzmann, 1898] by S.G. Brush, *Lecture on Gas Theory* Berkeley: University of California Press, 1964.

[Boltzmann, 1896b] L. Boltzmann. Entgegnung an die wärmetheoretischen Betrachtungen des Hrn. E., Zermelo. *Wiedemann's Annalen, 57*, 772–784, 1896. In [Boltzmann, 1909] Vol. III, paper 119.

[Boltzmann, 1897a] L. Boltzmann. Zu Hrn Zermelos Abhandlung "Über die mechanische Erklärung irreversibler Vorgänge". *Wiedemann's Annalen, 60* 392–398, 1897. in [Boltzmann, 1909], Vol. III paper 120.

[Boltzmann, 1897b] L. Boltzmann. Über einige meiner weniger bekannte Abhandlungen über Gastheorie und deren Verhältnis zu derselben *Verhandlungen des 69. Versammlung Deutscher Naturforscher und Ärzte*, Braunschweig 1897, pp. 19–26. *Jahresberichte der Deutsche Mathematiker-Vereinigung*, **6**, (1899), 130–138. Also in [Boltzmann, 1909] Vol. III, paper 123.

[Boltzmann, 1898] L. Boltzmann. *Vorlesungen über Gastheorie* Vol II. Leipzig, J.A. Barth, 1898. Translated, together with [Boltzmann, 1896] by S.G. Brush, *Lecture on Gas Theory* Berkeley: University of California Press, 1964.

[Boltzmann, 1898] L. Boltzmann. Über die sogenannte H-Kurve. *Mathematische Annalen* **50**, 325–332, 1898. In [Boltzmann, 1909] vol. III, paper 127.

[Boltzmann and Nabl, 1904] L. Boltzmann and J. Nabl. Kinetisch theorie der Materie. *Encyclopädie der Mathematischen Wisenschaften*, Vol V-1, pp. 493–557, 1904.

[Boltzmann, 1905] L. Boltzmann. *Populäre Schriften*. Leipzig: J.A. Barth, 1905. Re-issued Braunschweig: F. Vieweg, 1979.

[Boltzmann, 1909] L. Boltzmann. *Wissenschaftliche Abhandlungen* Vol. I, II, and III. F. Hasenöhrl (ed.) Leipzig, 1909. Reissued New York: Chelsea, 1969.

[Borel, 1914] E. Borel. *Le Hasard*. Paris: Alcan, 1914.

[Boyling, 1972] J. B. Boyling. An axiomatic approach to classical thermodynamics. *Proceedings of the Royal Society of London, 329*, 35–71, 1972.

[Bricmont, 1995] J. Bricmont. Science of chaos or chaos in science? *Physicalia, 17*, 159–208; also in P.R. Gross, N.Levitt and M.W.Lewis (Eds.),*The flight from science and reason*, New York: New York Academy of Sciences, pp. 131–176, 1995.

[Brown and Uffink, 2001] H. Brown and J. Uffink. (2001). The origins of time-asymmetry in thermodynamics: the minus first law. *Studies in History and Philosophy of Modern Physics, 32*, 525–538, 2001.

[Brush, 1966] S. G. Brush. *Kinetic theory* Vol. 1 and 2. Oxford, 1966: Pergamon.

[Brush, 1976] S. G. Brush. *The Kind of motion we call heat*. Amsterdam: North Holland, 1976.

[Brush, 1977] S. G. Brush. Statistical mechanics and the philosophy of science: some historical notes. In F. Supper & P.D. Asquith (Eds.) *PSA 1976 Proceedings of the Biennial meeting of the Philsopophy of Science Association 1976, Vol 2*, East Lansu=ing, Mich.: Philosophy of Science Association, pp. 551–584, 1976.

[Brush, 1999] S. G. Brush. Gadflies and geniuses in the history of gas theory. *Synthese 119*, 11–43. Also in [Brush, 2003], pp. 421–450, 1999.

[Brush, 2003] S. G. Brush. *The kinetic theory of gases*. London: Imperial College Press, 2003.

[Bryan, 1894] G. H. Bryan. Letter to the editor. *Nature, 51*, 175, 1894.

[Bryan, 1894] G. H. Bryan. The Assumption in Boltzmann's Minimum Theorem, *Nature, 52*, 29–30, 1894.

[Burbury, 1894a] S. H. Burbury. Boltzmann's minimum theorem. *Nature, 51*, 78–79, 1894.

[Burbury, 1894b] S. H. Burbury. The kinetic theory of gases. *Nature, 51* 175–176, 1894.

[Butterfield, 2006] J. N. Butterfield. This volume, chapter 1, 2006.

[Callender, 1999] C. Callender. Reducing thermodynamics to statistical mechanics: the case of entropy. *Journal of Philosophy, 96*, 348–373, 1999.

[Callender, 2001] C. Callender. Taking Thermodynamics Too Seriously. *Studies In History and Philosophy of Modern Physics, 32*, 539–553, 2001.

[Callender, 2004] C. Callender. Measures, explanations and the past: Should 'special' initial conditions be explained? *British Journal for Philosophy of Science, 55*, 195–217, 2004.

[Campisi, 2005] M. Campisi. On the mechanical foundations of thermodynamics: The generalized Helmholtz theorem *Studies In History and Philosophy of Modern Physics, 36*, 275–290, 2005.

[Casetti et al, 2003] L. Casetti, M. Pettini and E. G. D. Cohen. Phase Transitions and Topology Changes in Configuration Space. *Journal of Statistical Physics, 111* , 1091–1123, 2003.

[Carathéodory, 1909] C. Carathéodory. Untersuchungen über die Grundlagen der Thermodynamik, *Mathematische Annalen, 67*, 355–386, 1909. Translation by J. Kestin,"Investigation into the foundations of thermodynamics" in [Kestin, 1976], pp. 229–256. This translation is not quite accurate.

[Carnot, 1824] S. Carnot. *Réflexions sur la puissance motrice du feu.* Paris: Bachelier, 1824. Re-edited and translated in E. Mendoza (Ed.) Reflections on the motive power of fire, New York: Dover, 1960.

[Cercignani, 1998] C. Cercignani. *Ludwig Boltzmann, the man who trusted atoms* Oxford: Oxford University Press, 1998.

[Cercignani *et al.*, 1994] C. Cercignani, R. Illner, and M. Pulvirenti. *The mathematical theory of dilute gases.* New York: Springer-Verlag, 1994.

[Chang, 2003] H. Chang. Preservative realism and its discontents: revisiting caloric. *Philosophy of Science, 70*, 902–912, 2003.

[Chang, 2004] H. Chang. *Inventing temperature. Measurement and scientific progress.* Oxford: Oxford University Press, 2004.

[Clausius, 1857] R. Clausius Über die Art von Bewegung die wir Wärme nennen, *Poggendorff's Annalen, 100*, 253-280, 1857. English translation in [Brush, 1966], Vol.1. pp. 111–134.

[[, Clausius] 1862]Clausius62 R. Clausius. 'Ueber die Anwendung des Satzes von der Aequivalenz der Verwandlungen auf die innere Arbeit', *Viertelsjahrschrift der Züricher naturforschenden Gesellschaft, 7*, 48, 1862. Also in *Annalen der Physik und Chemie, 116*: 73–112 (1862), English translation in *Philosophical Magazine, 24*, 81–97, 201–213, also in [Kestin, 1976], pp. 133–161.

[Clausius, 1865] R. Clausius. Über verschiedene für die Anwendung bequeme Formen der Haubtgleichungen der mechanische Wärmetheorie, *Poggendorff's Annalen, 100*, (253), 1865. English translation in [Kestin, 1976].

[Cohen, 1996] E. G. D. Cohen. Boltzmann and statistical mechanics, cond-mat/9608054v2, 1996.

[Compagner, 1989] A. Compagner. Thermodynamics as the continuum limit of statistical mechanics *American Journal of Physics, 57*, 106–117, 1989.

[Cornfeld *et al.*, 1982] I. P. Cornfeld, S. V. Fomin, and Ya. G. Sinai. *Ergodic theory* New York: Springer-Verlag, 1982.

[Culverwell, 1894] E. Culverwell. Dr. Watson's proof of Boltzmann's theorem on permanence of distributions *Nature, 50* 617, 1894.

[Culverwell, 1895] E. Culverwell. Boltzmann's minimum theorem *Nature, 51* 246, 1895.

[Curd, 1982] M. Curd. Popper on the direction of time. In R. Sexl and J. Blackmore (eds.), *Ludwig Boltzmann, internationale Tagung: anlässlich des 75. Jahrestages seines Todes, 5.-8. September 1981: ausgewählte Abhandlungen*, pp. 263-303, 1982. Graz: Akademische Druck- und Verlagsanstalt.

[Davies, 1974] E. B. Davies. Markovian master equations. *Communications in Mathematical Physics, 39* 91-110, 1974.

[Davies, 1976a] E. B. Davies. Markovian master equations II *Mathematische Annalen, 219* 147–158, 1976.

[Davies, 1976] E. B. Davies. *Quantum theory of open systems.* New York: Academic Press, 1976.

[Denbigh and Denbigh, 1985] K. G. Denbigh and J. Denbigh. *Entropy in relation to incomplete knowledge.* Cambridge: Cambridge University Press, 1985.

[Dias, 1994] P. M. C. Dias. Will someone say exactly what the H-theorem proves? A study of Burbury's Condition A and Maxwell's Proposition II. *Archive for History of Exact Sciences 46*, 341–366, 1994.

[Dugas, 1959] R. Dugas. *La théorie physique au sens de Boltzmann et ses prolongements modernes.* (Neuchâtel: Griffon), 1959.

[Earman, 2006] J. Earman. The past-hypothesis: not even false. *Studies in History and Philosophy of Modern Physics*, 2006.

[Earman and Rédei, 1996] J. Earman and M. Rédei. Why ergodic theory does not explain the success of equilibrium statistical mechanics. *British Journal for the Philosophy of Science, 45* 63–78, 1996.

[Earman and Norton, 1998] J. Earman and J. D. Norton. Exorcist XIV: The wrath of Maxwell's demon. Part I. From Maxwell to Szilard. *Studies in History and Philosophy of Modern Physics, 29*, 435–471, 1998.

[Earman and Norton, 1999] J. Earman and J. D. Norton. Exorcist XIV: The wrath of Maxwell's demon. Part II. From Szilard to Landauer and beyond *Studies in History and Philosophy of Modern Physics, 30*, 1–40.

[Edens, 2001] B. Edens. Semigroups and symmetry: an investigation of Prigogine's theories. http://philsci-archive.pitt.edu/archive/00000436/, 2001.

[Ehrenfest and Ehrenfest-Afanassjewa, 1912] P. Ehrenfest and T. Ehrenfest-Afanassjewa. Begriffliche Grundlagen der Statistischen Auffassung in der Mechanik. *Enzyclopädie der Mathematischen Wissenschaften* Vol. 4. F. Klein and C. Müller (eds.). Leibzig: Teubner, pp. 3–90, 1912. English translation *The conceptual foundations of the statistical approach in mechanics.* Ithaca N.Y.: Cornell University Press, 1959.

[Einstein, 1902] A. Einstein. Kinetische Theorie des Wärmegleichgewicht und des zweiten Hauptstzes der Thermodynamik. *Annalen der Physik, 9*, 417–433, 1902.

[Einstein, 1910] A. Einstein. Theorie der Opaleszenz von homogenen Flüssigkeiten und Flüssigkeitsgemischen in der Nähe des kritischen Zustandes *Annalen der Physik, 33*, 1275–1298, 1910.

[Ellis, 1850] R. L. Ellis. Remarks on an alleged proof of the method of least squares, contained in a late number of the Edinburgh Review, in W.Walton (Ed.), Mathematical and other Writings of R.L. Ellis. Cambridge: Cambridge University Press, 1863, pp. 53–61.

[Emch, 1965] G. G. Emch. On the Markov character of master-equations. *Helvetica Physica Acta, 38*, 164–171, 1965.

[Emch, 2005] G. G. Emch. Probabilistic issues in statistical mechanics. *Studies In History and Philosophy of Modern Physics 36*, 303–322, 2005.

[Emch, 2006] G. G. Emch. This volume, chapter 10, 2006.

[Emch and Liu, 2001] G. G. Emch and C. Liu, *The Logic of thermostatistical physics.* Berlin: Springer, 2001.

[Farquhar, 1964] I. E. Farquhar. *Ergodic theory in statistical mechanics*, London, Interscience, 1964.

[Feshbach, 1987] H. Feshbach. Small systems: when does thermodynamics apply? *Physics Today, 40*, 9–11, 1987.

[Fine, 1973] T. L. Fine. *Theories of probability: An examination of foundations.* New York: Academic Press, 1973.

[Fisher, 1964] M. E. Fisher. The free energy of a macroscopic system. *Archive for Rational Mechanics and Analysis 17* 377–410, 1964.

[Fowler and Guggenheim, 1939] R. H. Fowler and E. Guggenheim. *Statistical Thermodynamics.* Cambridge: Cambridge University Press, 1939.

[Fox, 1971] R. Fox. *The caloric theory of gases: from Lavoisier to Regnault.* Oxford: Clarendon Press, 1971.

[Friedman, 1976] K. S. Friedman. A Partial vindication of ergodic theory. *Philosophy of Science, 43*, 151-162, 1976.

[Frigg, 2004] R. Frigg. In what sense is the Kolmogorov-Sinai entropy a measure for chaotic behaviour? — bridging the gap between dynamical systems theory and communication theory. *British Journal for the Philosophy of Science, 55*, 411–434, 2004.

[Galavotti, 2004] M. C. Galavotti. *A philosophical introduction to probability.* Chicago: University of Chicago Press, 2004.

[Gallavotti, 1994] G. Gallavotti. Ergodicity, ensembles, irreversibility in Boltzmann and beyond http: arXiv:chao-dyn/9403004. *Journal of Statistical Physics, 78*, 1571–1589, 1994.

[Gallavotti, 1999] G. Gallavotti. *Statistical mechanics: A short treatise.* Berlin: Springer, 1999.

[Gantmacher, 1959] F. R. Gantmacher. *Matrizenrechnung* Vol 2. Berlin: Deutscher Verlag der Wissenschaften, 1959.

[Garber *et al.*, 1986] E. Garber, S. G. Brush, and C. W. F. Everitt, (Eds.). *Maxwell on molecules and gases* Cambridge Mass.: MIT Press, 1986.

[Garber *et al.*, 1995] E. Garber, S. G. Brush, and C. W. F. Everitt, (Eds.). *Maxwell on heat and statistical mechanics* Bethlehem: Lehigh University Press, 1995.

[Gearhart, 1990] C. A. Gearhart. Einstein before 1905: The early papers on statistical mechanics *American Journal of Physics, 58*, 468–480, 1990.

[Gibbs, 1875] J. W. Gibbs. equilibrium of heterogenous substances. *Transactions of the Connecticut Academy, 3*, 103–246 and 343–524 (1878). Also in [Gibbs, 1906, pp. 55–353].

[Gibbs, 1902] J. W. Gibbs. *Elementary principles in statistical mechanics*, New York, Scribner etc, 1902.

[Gibbs, 1906] J. W. Gibbs. *The Scientific Papers of J. Willard Gibbs*, Vol. 1, Thermodynamics, Longmans, London, 1906. Reissued by Ox Bow Press, Woodbridge, Connecticut, 1993.

[Giles, 1964] R. Giles. *Mathematical foundations of thermodynamics* Oxford: Pergamon, 1964.

[Gold, 1956] T. Gold. Cosmic processes and the nature of time. In R. Colodny (Ed.), *Mind and Cosmos.* Pittsburgh: University of Pittsburgh Press, pp. 311–329, 1956.

[Goldstein, 2001] S. Goldstein. Boltzmann's approach to statistical mechanics. In J. Bricmont, D. Dürr, M.C. Galavotti, G. Ghirardi, F. Petruccione, and N. Zanghi (Eds.) *Chance in physics: foundations and perspectives*, Lecture Notes in Physics 574. Berlin: Springer-Verlag, pp. 39–54, 2001. Also as e-print cond-mat/0105242.

[Gorini et al., 1976] V. Gorini, A. Kossakowski, and E. C. G. Sudarshan. Completely positive dynamical semigroups of N-level systems. *Journal of Mathematical Physics 17* 8721–825, 1976.

[Grimmett and Stirzaker, 1982] G. R. Grimmett and D. R. Stirzaker. *Probability and random processes.* Oxford: Clarendon Press, 1982.

[de Groot, 1951] S. de Groot. *Thermodynamics of irreversible processes.* Amsterdam: North-Holland, 1951.

[de Groot and Mazur, 1961] S. R. de Groot and P. Mazur. *Non-equilibrium thermodynamics.* Amsterdam: North-Holland, 1961.

[Gross, 1997] D. H. E. Gross. Microcanonical thermodynamics and statistical fragmentation of dissipative systems. The topological structure of the N-body phase space. *Physics Reports, 279*, 119–201, 1997.

[Gross and Votyakov, 2000] D. H. E. Gross and E. V. Votyakov. Phase transitions in "small" systems. *European Physical Journal B 15*, 115–126, 2000.

[Grünbaum, 1973] A. Grünbaum. Is the coarse-grained entropy of classical statistical mechanics an anthropomorphism? In *Philosophical problems of space and time*, 2nd ed., Dordrecht: Reidel, pp. 646–665, 1973.

[Guthrie, 1874] F. Guthrie. Molecular motion. *Nature, 10* 123, 19=874. Also in [Garber *et al.*, 1995, p.143–145].

[Greven et al., 2003] A. Greven, G. Keller, and G. Warnecke, (Eds.) *Entropy.* Princeton: Princeton University Press, 2003.

[ter Haar, 1955] D. ter Haar. Foundations of statistical mechanics. *Reviews of Modern Physics, 27*, 289–338, 1955.

[Hacking, 1975] I. Hacking. *The emergence of probability.* Cambridge: Cambridge University Press, 1975.

[Hacking, 1990] I. Hacking. *The taming of Chance.* Cambridge: Cambridhge University Press, 1990.

[Van Harn and Holewijn, 1991] K. van Harn and P. J. Holewijn. *Markov-ketens in diskrete tijd.* Utrecht: Epsilon, 1991.

[Herschel, 1850] J. F. W. Herschel. Quetelet on Probabilities *Edinburgh Review.* Also in *Essays from the Edinburgh and Quarterly Reviews with addresses and other pieces*, London: Longman, Brown, Green, Longmans and Roberts, 1857, pp. 365–465, 1850.

[Hertz, 1910] P. Hertz. Über die mechanischen Grundlagen der Thermodynamik. *Annalen der Physik, 33*, 225–274; 537–552, 1910.

[Hill, 1987] T. L. Hill. *Thermodynamics of small systems.* New York: Benjamin, 1987.

[Höflechner, 1994] W. Höflechner. *Ludwig Boltzmann Leben und Briefe.* Graz: Akademische Druck- und Verlagsanstalt, 1994.

[Huang, 1987] K. Huang. *Statistical mechanics.* New York: Wiley, 1987.

[Huggett, 1999] N. Huggett. Atomic metaphysics. *Journal of Philosophy, 96*, 5–24, 1999.

[Hutter and Wang, 2003] K. Hutter and Y. Wang. Phenomenological thermodynamics and entropy principles, 2002. In [Greven et al., 2003, pp. 55–78].

[Illner and Neunzert, 1987] R. Illner and H. Neunzert. The concept of irreversibility in kinetic theory of gases. *Transport Theory and Statistical Physics, 16*, 89–112, 1987.

[Janssen, 2002] M. Janssen. Dog fleas and tree trunks: the Ehrenfests marking the territory of Boltzmann's H-theorem. Unpublished, 2002.

[Jauch, 1972] J. Jauch. On a new foundation of equilibrium thermodynamics. *Foundations of Physics, 2*, 327–332, 1972.

[Jauch, 1975] J. Jauch. Analytical thermodynamics. Part 1. Thermostatics–general theory. *Foundations of Physics, 5*, 111–132, 1975.

[Jaynes, 1965] E. T. Jaynes. Gibbs vs. Boltzmann entropies. *American Journal of Physics, 33*, 391–398, 1965. Also in [Jaynes, 1983, pp. 77–86].

[Jaynes, 1967] E. T. Jaynes. Foundations of probability theory and statistical mechanics. In: M. Bunge (Ed.) *Delaware seminar in the foundations of physics*. Berlin: Springer-Verlag, pp. 77–101, 1967. Also in [Jaynes, 1983, pp. 89–113].

[Jaynes, 1983] E. T. Jaynes. *Probability, statistics and statistical physics*. R. Rosenkrantz (Ed.) Dordrecht: Reidel, 1983.

[Jaynes, 1992] E. T. Jaynes. The Gibbs paradox. In C.R. Smith, G.J. Erickson, & P.O. Neudorfer, (Eds.) *Maximum entropy and Bayesian methods*. Dordrecht: Kluwer, pp. 1–22, 1992.

[Jepps *et al.*, 2000] O. G. Jepps, G. Ayton, and D. J. Evans. Microscopic expressions for the thermodynamic temperature. *Physical Review E, 62*, 4757–4763, 2000.

[Kadanoff, 2000] L. P. Kadanoff. *Statistical physics : statics, dynamics and renormalization* Singapore: World Scientific, 2000.

[van Kampen, 1962] N. G. van Kampen. Fundamental problems in the statistical mechanics of irreversible processes. In E.G.D. cohen (Ed.), *Fundamental problems in statistical mechanics* Amsterdam: North-Holland, pp.173–202, 1962.

[van Kampen, 1981] N. G. van Kampen. *Stochastic processes in chemistry and physics*. Amsterdam: North-Holland, 1981.

[van Kampen, 1984] N. G. van Kampen. The Gibbs paradox. In W.A. Parry (ed.), *Essays in theoretical physics in honour of Dirk ter Haar*, Oxford: Pergamon Press, pp. 303–312, 1984.

[van Kampen, 1994] N. G. van Kampen. Models for dissitpation in quantumn mechanics. In J.J.Brey, J. Marro, J.M. Rubi, M. San Miguel (Eds.) *25 years of non-equilibrium statistical mechanics* Berlin: Springer, 1994.

[van Kampen, 2002] N. G. van Kampen. The road from molecules to Onsager. *Journal of Statistical Physics, 109*, 471–481, 2002.

[van Kampen, 2004] N. G. van Kampen. A new approach to noise in quantum mechanics *Journal of Statistical Physics, 115*, 1057–1072, 2004.

[Karakostas, 1996] V. Karakostas. On the Brussels school's arrow of time in quantum theory. *Philosophy of Science, 63*, 374–400, 1996.

[Kelly, 1979] F. P. Kelly. *Reversibility and stochastic networks* Chichester: Wiley, 1979. Also at http://www.statslab.cam.ac.uk/ afrb2/kelly_book.html.

[Kestin, 1976] J. Kestin. *The second law of thermodynamics*. Stroudsburg, Pennsylvania: Dowden, Hutchinson and Ross, 1976.

[Keynes, 1921] J. M. Keynes. *A treatise on probability*. London: Macmillan, 1921.

[Khinchin, 1949] A. Khinchin. *Mathematical foundations of statistical mechanics*. (New York: Dover), 1949.

[Kirchhoff, 1894] G. Kirchhoff. *Vorlesungen über mathematische Physik, Vol IV: Theorie der Wärme* M. Planck (Ed.). Leipzig: Teubner, 1894.

[Kirsten and Körber, 1975] C. Kirsten and H.-G. Körber. *Physiker über Physiker*, Berlin: Akademie-Verlag, 1975.

[Klein, 1970] M. J. Klein. Maxwell, his demon and the second law of thermodynamics, *American Scientist 58*, 82–95, 1970; also in [Leff and Rex, 1987, pp. 59–72].

[Klein, 1972] M. J. Klein. Mechanical explanation at the end of the nineteenth century. *Centaurus, 17*, 58–82,1972.

[Klein, 1973] M. J. Klein. The Development of Boltzmann's Statistical Ideas. In E.G.D. Cohen & W. Thirring (eds.), *The Boltzmann equation*, Wien: Springer, pp. 53–106, 1973.

[Klein, 1974] M. J. Klein. Boltzmann, monocycles and mechanical explanation. In R.J. Seeger & R.S. Cohen (Eds.), *Philsophical foundations of science; Boston Studies in the Philosophy of Science XI* Dordrecht: Reidel, pp. 155–175, 1974.

[Klein, 1978] M. J. Klein. The thermostatics of J. Willard Gibbs: atyransformation of thermoidynamics. In E.G. Forbes (Ed.) *Human implications of scientific advance*. Edinburgh: Edinburgh University Press, pp. 314–330, 1978.

[Kroes, 1985] P. A. Kroes. *Time: its structure and role in physical theories.* Dordrecht: Reidel, 1985.

[Kurth, 1960] R. Kurth. Axiomatics of classical statistical mechanics. Oxford: Pergamon Press, 1960.

[Ladyman et al., 2006] J. Ladyman, S. Presnell, T. Short, and B. Groisman. The connection between logical and thermodynamical irreversibility. http://philsci-archive.pitt.edu/archive/00002374/.

[Landau and Lifshitz, 1987] L. E. Landau and E. M. Lifshitz. *Fluid Mechanics* 3rd Edition, Oxford: Butterworth-Heinemann, 1987.

[Lanford, 1973] O. E. Lanford. Entropy and equilibrium states in classical statistical mechancs. In A. Lenard (Ed.) *Statistical mechanics and mathematical problems.* Berlin: Springer-Verlag, pp. 1–113.

[Lanford, 1975] O. E. Lanford. Time evolution of large classical systems. In J. Moser (Ed.) *Dynamical Systems, Theory and Applications*, Lecture Notes in Theoretical Physics Vol. 38, Berlin: Springer, pp. 1–111, 1975.

[Lanford, 1976] O. E. Lanford. On a derivation of the Boltzmann equation. *Astérisque, 40* 117–137, 1976. Also in J.L Lebowitz & E.W. Montroll (Eds.) *Nonequilibrium phenomena I: the Boltzmann Equation.* Amsterdam: North-Holland, 1983.

[Lanford, 1981] O. E. Lanford. The hard sphere gas in the Boltzmann-Grad limit. *Physica, 106A*, 70–76, 1981.

[Lavis, 2004] D. A. Lavis. The spin-echo system reconsidered. *Foundations of Physics, 34*, 669–688, 2004.

[Lavis, 2005] D. A. Lavis. Boltzmann and Gibbs: An attempted reconciliation. *Studies In History and Philosophy of Modern Physics, 36*, 245–273, 2005.

[Lebowitz, 1994] J. Lebowitz. Time's arrow and Boltzmann's entropy. In J.J. Halliwell, J. Pérez-Mercader & W.H. Zurek (Eds.). *Physical origins of time asymmetry.* Cambridge: Cambridge University Press, pp. 131–146, 1994.

[Lebowitz, 1999] J. L. Lebowitz. Statistical mechanics: A selective review of two central issues. *Reviews of Modern Physics, 71*, S346–S357, 1999. math-ph/0010018.

[Lebowitz and Penrose, 1973] J. Lebowitz and O. Penrose. Modern ergodic theory. *Physics Today, 26* 23–29, 1973.

[Leeds, 1989] S. Leeds. Discussion: D. Malament and S. Zabell on Gibbs phase averaging. *Philosophy of Science, 56*, 325–340, 1989.

[Leff and Rex, 1987] H. S. Leff and A. F. Rex. *Maxwell's Demon 2.* Bristol: Institute of Physics, 1987.

[Leinaas and Myrheim, 1977] J. M. Leinaas and J. Myrheim. On the theory of identical particles. *Il Nuovo Cimento, 37B*, 1–23, 1977.

[Lieb, 1976] E. H. Lieb. The stability of matter. *Reviews of Modern Physics, 48*, 553–569, 1976.

[Lieb and Lebowitz, 1973] E. H. Lieb and J. L. Lebowitz. Lectures on the thermodynamic limit for Coulomb systems. In A. Lenard (Ed.) *Statistical mechanics and mathematical problems.* Berlin: Springer-Verlag, pp. 136–162, 1973.

[Lieb and Yngvason, 1999] E. Lieb and J. Yngvason. The phyics and mathematics of the second law of thermodynamics, *Physics Reports, 310*, 1–96; erratum 314, (1999), p. 669, 1999. Also as e-print: http://xxx.lanl.gov/abs/cond-mat/9708200.

[Lindblad, 1976] G. Lindblad. On the generators of quantum dynamical semigroups. *Communications in Mathematical Physics, 48*, 119–130, 1976.

[Lindblad, 1983] G. Lindblad. *Non-equilibrium entropy and irreversibility* Dordrecht: Reidel, 1983.

[van Lith, 2001a] J. van Lith. Ergodic theory, interpretations of probability and the foundations of statistical mechanics *Studies in History and Philosophy of Modern Physics, 32*, 581–594, 2001.

[van Lith, 2001b] J. van Lith. *Stirr in Stillness.* Ph.D. Thesis, Utrecht University, 2001.

[Liu, 1999] C. Liu. Explaining the emergence of cooperative phenomena. *Philosophy of Science 66 (Proceedings)*, S92–S106, 1999.

[Lorentz, 1887] H. A. Lorentz. Über das Gleichgewicht der lebendige Kraft unter Gasmolekülen. *Wiener Berichte, 95*, 115–152, 187, 1887. Also in: *Collected Papers* The Hague: Martinus Nijhoff 1938, pp. 74–111.

[Lorentz, 1916] H. A. Lorentz. *Les theories statistiques en thermodynamique*. Leipzig: Teubner, 1916.

[Loschmidt, 1876] J. Loschmidt. Über die Zustand des Wärmegleichgewichtes eines Systems von Körpern mit Rücksicht auf die Schwerkraft. *Wiener Berichte, 73*, 128–142, 366–372 (1876), 75, 287–298, 76, 209–225 (1877).

[Maes and Netočný, 2003] C. Maes and K. Netočný. Time-reversal and entropy. *Journal of Statistical Physics, 110*, 269–310, 2003.

[Mackey, 1992] M. C. Mackey. *Time's arrow: the origins of thermodynamic behavior*. New York: Springer, 1992.

[Mackey, 2001] M. C. Mackey. Microscopic dynamics and the second law of thermodynamics. In: *Time's Arrows, quantum Measurements and Superluminal Behavior*. C. Mugnai, A. Ranfagni & L.S. Schulman (Eds.) (Roma: Consiglio Nazionale delle Ricerche), 2001.

[Malament and Zabell, 1980] D. B. Malament and S. L. Zabell. Why Gibbs Phase Averages Work — The Role of Ergodic Theory. *Philosophy of Science, 56*, 339–349, 1980.

[Mandelbrot, 1956] B. Mandelbrot. An outline of a purely phenomenological theory of statistical thermodynamics: 1. canonical ensembles. *IRE Transactions on Information Theory, IT-2*, 190–203, 1956.

[Mandelbrot, 1962] B. Mandelbrot. The role of sufficiency and of estimation in thermodynamics. *Annals of Mathematical Statistics, 33*, 1021–1038, 1962.

[Mandelbrot, 1964] B. Mandelbrot. On the derivation of statistical thermodynamics from purely phenomenological principles. *Journal of Mathematical Physics, 5*, 164–171. 1964.

[Mañé, 1987] R. Mañé. *Ergodic theory and differentiable dynamics*. Berlin: Springer, 1987.

[Maroney, 2005] O. J. E. Maroney. The (absence of a) relationship between thermodynamic and logical reversibility. *Studies In History and Philosophy of Modern Physics, 36*, 355–374, 2005.

[Martin-Löf, 1979] A. Martin-Löf. *Statistical mechanics and the foundations of thermodynamics*. Berlin: Springer-Verlag, 1979.

[Maxwell, 1860] J. C. Maxwell. Illustrations of the dynamical theory of gases. *Philosophical Magazine, 19*, 19–32; 20, 21–37, 19=860. Also in [Garber *et al.*, 1986, pp. 285–318].

[Maxwell, 1867] J. C. Maxwell. On the dynamical theory of gases. *Philosophical Transactions of the Royal Society of London, 157* 49-88 (1867). Also in [Brush, 1966] and [Garber *et al.*, 1986, pp. 419–472]

[Maxwell, 1872] J. C. Maxwell. *Theory of heat.*(2nd ed.) London: Longmans, Green and Co, 1872.

[Maxwell, 1873] J. C. Maxwell. Molecules. *Nature, 8*, 437–441 (1873). Also in [Garber *et al.*, 1986, pp. 138–154]

[Maxwell, 1873] J. C. Maxwell. On the final state of a system of molecules in motion subject to forces of any kind. *Nature, 8*, 537-538 (1867), 1873. Also in [Garber *et al.*, 1995, pp. 138–143]

[Maxwell, 1875] J. C. Maxwell. On the dynamical evidence of the molecular constitution of bodies. *Nature, 11* 357-359, 374–377, (1875). Also in [Garber *et al.*, 1986, pp. 216–237].

[Maxwell, 1877] J. C. Maxwell. A treatise on the kinetic theory of gases by Henry William Watson, *Nature, 18*, 242–246, 1877. Also in [Garber *et al.*, 1995, p. 156–167].

[Maxwell, 1878a] J. C. Maxwell. Constitution of bodies. *Encyclopaedia Brittanica*, ninth edition, Vol. 6, pp. 310–313, 1878. Also in [Garber *et al.*, 1986, pp. 246–254].

[Maxwell, 1878b] J. C. Maxwell. Diffusion. *Encylopedia Brittanica* ninth edition, Vol. 7, pp. 214–221, 1978; also in [Garber *et al.*, 1986, pp. 525–546].

[Maxwell, 1878c] J. C. Maxwell. On stresses in rarified gases arising from inequalities of temperature *Proceedings of the Royal Society of London, 27* 304–308; *Philosophical Transactions of the Royal Society of London, 170*, (1880) 231–256, 1978. Also in [Garber *et al.*, 1995, pp. 357–386].

[Maxwell, 1879] J. C. Maxwell. On Boltzmann's theorem on the average distribution of energy in a system of material points. *Transactions of the Cambridge Philosophical Society, 12*, 547–570, 1879. Also in [Garber *et al.*, 1995, pp. 357–386].

[Mehra, 1998] J. Mehra. Josiah Willard Gibbs and the Foundations of Statistical Mechanics *Foundations of Physics, 28*, 1785–1815, 1998.

[Mehra and Sudarshan, 1972] J. Mehra and E. C. G. Sudarshan. Some reflections on the nature of entropy, irreversibility and the second law of thermodynamics, *Nuovo Cimento B, 11*, 215–256, 1972.

[Meixner, 1969] J. Meixner. Processes in simple thermodynamic materials *Archive for Rational Mechanics and Analysis 33*, 33–53, 1969.

[Meixner, 1970] J. Meixner. On the foundations of thermodynamics of processes. In B. Gal-Or, E.B. Stuart & A. Brainard (Eds.), *A Critical Review of Thermodynamics*. Baltimore: Mono Book Corporation, pp. 37–47, 1970. Also in [Kestin, 1976], pp. 313–323.

[Moran, 1961] P. A. P. Moran. Entropy, Markov processes and Boltzmann's H-theorem. *Proceedings of the Cambridge Philosophical Society, 57*, 833–842, 1961.

[Morrison, 2000] M. Morrison. *Unifying scientific theories: physical concepts and mathematical structures*. Cambridge: Cambridge University Press. 2000.

[Morrison, 1966] P. Morrison. Time's arrow and external perturbations. In A. de Shalit, H. Feshbach, & L. van Hove (Eds.), *Preludes in Theoretical Physics in honor of V. F. Weisskopf*. Amsterdam: North Holland, pp. 347–, 1966.

[Müller, 2003] I. Müller. Entropy in non-equilibrium. In [Greven et al., 2003, pp. 79–107].

[Navarro, 1998] L. Navarro. Gibbs, Einstein and the foundations of statistical mechanics. *Archive for History of Exact Science, 53*, 147–180, 1998.

[Nemytskii and Stepanov, 1960] Nemytskii, V.V., & V. V. Stepanov. *Qualitative theory of differential equations*. Princeton: Princeton University Press, 1960.

[Norton, 2005] J. D. Norton. Eaters of the lotus: Landauer's principle and the return of Maxwell's demon. *Studies In History and Philosophy of Modern Physics, 36*, 375–411, 2005.

[von Neumann, 1932] J. von Neumann. Proof of the quasi-ergodic hypothesis. *Proceedings of the National Academy of sciences of the United States of America, 18*, 70–82 and 263–266, 1932.

[Obcemea and Brändas, 1983] Ch. Obcemea and E. Brändas. Analysis of Prigogine's theory of subdynamics. *Annals of Physics, 147*, 383–430, 1983.

[Olsen, 1993] E. T. Olsen. Classical mechanics and entropy. *Foundations of Physics Letters, 6*, 327–337, 1993.

[Penrose, 1970] O. Penrose. *Foundations of statistical mechanics: a deductive treatment*. Oxford: Pergamon Press, 1970.

[Penrose, 1979] O. Penrose. Foundations of statistical mechanics. *Reports on Progress in Physics, 42*, 1937–2006, 1979.

[Penrose and Percival, 1962] O. Penrose and I. Percival. The direction of time. *Proceedings of the Physical Society, 79*, 605–616, 1962.

[Pešić, 1991] P. D. Pešić. The principle of identicality and the foundations of quantum theory. I. The Gibbs paradox. *American Journal of Physics, 59*, 971–974, 1991.

[Petersen, 1983] K. Petersen. *Ergodic theory*. Cambridge: Cambridge University Press, 1983.

[Pitowsky, 2001] I. Pitowsky. Local fluctuations and local observers in equilibrium statistical mechanics. *Studies In History and Philosophy of Modern Physics, 32*, 595–607, 2001.

[Plancherel, 1913] M. Plancherel. Foundations of statistical mechanics. *Annalen der Physik, 42*, 1061–1063, 1913.

[von Plato, 1991] J. von Plato. Boltzmann's ergodic hypothesis. *Archive for History of Exact Sciences, 42*, 71–89, 1991.

[von Plato, 1994] J. von Plato. *Creating modern probability*. Cambridge: Cambridge University Press, 1994.

[Poincaré, 1889] H. Poincaré. Sur les tentatives d'explication mécanique des principes de la thermodynamique. *Comptes Rendus de l'Académie des Sciences (Paris), 108*, 550–553, 1889. English translation in [Olsen, 1993].

[Poincaré, 1893] H. Poincaré. Le mécanisme et l'expérience. *Revue de Métaphysique et de Morale, 1*, 534–537, 1893. English translation in [Brush, 1966].

[Poincaré, 1896] H. Poincaré. *Calcul des Probabilités*. Paris: Carré, 1896.

[Popper, 1982] K. Popper. *Quantum theory and the schism in physics*. London: Hutschinson, 1982.

[Price, 1996] H. Price. *Time's arrow and Archimedes' point*. New York: Oxford University Press, 1996.

[Prigogine, 1955] I. Prigogine. *Introduction to the thermodynamics of irreversible processes*. New York: Interscience, 1955.

[Ray, 1984] J. R. Ray. Correct Boltzmann counting. *European Journal of Physics, 5* 219–224, 1984.

[Redhead, 1995] M. Redhead. *From physics to metaphysics*. Cambridge: Cambridge University Press, 1995.

[Reichenbach, 1956] H. Reichenbach. *The direction of Time*. Berkeley: University of California Press, 1956.

[Ridderbos, 2002] T. M. Ridderbos. The coarse-graining approach to statistical mechanics: how blissful is our ignorance? *Studies in History and Philosophy of Modern Physics, 33*, 65–77, 2002.

[Ridderbos and Redhead, 1998] T. M. Ridderbos and M. L. G. Redhead. The spin-echo experiment and the second law of thermodynamics. *Foundations of Physics, 28*, 1237–1270, 1998.

[Rosenthal, 1913] A. Rosenthal. Beweis der Unmöglichkeit ergodischer mechanischer Systeme. *Annalen der Physik, 42*, 796–806, 1913. English translation in [Brush, 2003, p.505–523].

[Rovelli, 2006] C. Rovelli. This volume, chapter 12, 2006.

[Ruelle, 1969] D. Ruelle. *Statistical mechanics: rigorous results*. New York: Benjamin, 1969.

[Rugh, 1997] H. H. Rugh. Dynamical approach to temperature *Physical Review Letters 78*, 772–774, 1997.

[Rugh, 2001] H. H. Rugh. Microthermodynamic formalism *Physical Review E 64*, 055101, 2001.

[Saunders, 2006] S. Saunders. On the explanation for quantum statistics. *Studies in History and Philosophy of Modern Physics, 37*, 2006.

[Schrödinger, 1950] E. Schrödinger. Irreversibility. *Proceedings of the Royal Irish Academy 53*, 189–195, 1950.

[Simányi and Szász, 1999] N. Simányi and D. Szász. Hard balls systems are completely hyperbolic. *Annals of Mathematics, 149*, 35–96, 1999.

[Sinai, 1963] Ya. G. Sinai. On the foundation of the ergodic hypothesis for a dynamical system of statistical mechanics *Soviet Mathematics Doklady, 4*, 1818–1822, 1963.

[Sinai and Chernov, 1987] Ya. G. Sinai and N. I. Chernov. Ergodic properties of certain systems of $2-D$ discs and $3-D$ balls. *Russian Mathematical Surveys, 42* 181–207. Also in Ya.G. Sinai (Ed.) *Dynamical systems; collection of papers* Singapore: Wold Scientific (1991) pp. 345–371, 1987.

[Shenker, 2000] O. Shenker. Interventionism in statistical mechanics: some philosophical remarks. http://philsci-archive.pitt.edu/archive/00000151/ 2000.

[Sklar, 1973] L. Sklar. Statistical explanation and ergodic theory. *Philosophy of Science, 40*, 194–212, 1973.

[Sklar, 1993] L. Sklar. *Physics and Chance. Philosophical Issues in the Foundations of Statistical Mechanics*. Cambridge: Cambridge University Press, 1993.

[Sklar, 2002] L. Sklar. *Theory and thruth*. Oxford: Oxford University Press, 2002.

[Spohn, 1980] H. Spohn. Kinetic equations from Hamiltonian dynamics: Markovian limits. *Reviews of Modern Physics 52*, 569–615, 1980.

[Spohn, 1991] H. Spohn. *Large Scale Dynamics of interacting Particles*. Berlin: Springer, 1991.

[Streater, 1995] R. F. Streater. *Statistical dynamics; a stochastic approach to non-equilibrium thermodynamics*. London: Imperial College Press, 1995.

[Styer, 2004] D. F. Styer. What good is the thermodynamic limit?" *American Journal of Physics 72*, 25–29; Erratum p. 1110, 2004.

[Sudarshan et al., 1961] E. C. G. Sudarshan, P. M. Mathews, and J. Rau. Stochastic dynamics of quantum-mechanical systems, *Physical Review, 121*, 920–924, 1961.

[Szász, 1996] D. Szász. Boltzmann's ergodic hypothesis, a conjecture for centuries? *Studia Scientiarum Mathematicarum Hungaria, 31*, 299–322, 1996.

[Szilard, 1925] L. Szilard. Über die Ausdehnung der phänomenologischen Thermodynamik auf die Schwankungserscheinungen. *Zeitschrift für Physik, 32*, 753–788, 1925.

[Tabor, 1989] M. Tabor. *Chaos and integrability in nonlinear dynamics: an introduction*. New York: Wiley, 1989.

[Thomsen and Hartka, 1962] J. S. Thomsen and Th. J. Hartka. Strange Carnot cycles; thermodynamics of a system with a density extremum *American Journal of Physics 30*, 26–33, 1962.

[Thompson, 1972] C. Thompson. *Mathematical Statistical Mechanics*. Princeton: Princeton University Press, 1972.

[Tisza, 1966] L. Tisza. *Generalized Thermodynamics*. Cambridge, Mass.: MIT Press, 1966.

[Tisza and Quay, 1963] L. Tisza and P. M. Quay. The statistical thermodynamics of equilibrium. *Annals of Physics, 25*, 48–90, 1963. Also in [Tisza, 1966, pp. 245–287].

[Tolman, 1938] R. C. Tolman. *The principles of statistical mechanics*. London: Oxford University Press, 1938.

[Touchette, 2003] H. Touchette. *Equivalence and nonequivalence of the microcanonical and canonical ensembles: a large deviations study*. Ph.D Thesis, McGill University, Montréal, 2003.

[Touchette et al., 2004] H. Touchette, R. S. Ellis, and B. Turkington. An introduction to the thermodynamic and macrostate levels of nonequivalent ensembles. *Physica A 340*, 138-146, 2004.

[Truesdell, 1961] C. Truesdell. Ergodic theory in classical statistical mechanics. In P. Caldirola (ed.), *Ergodic theories*, New York: Academic Press, 1961.

[Truesdell, 1969] C. Truesdell. *Rational thermodynamics.* New York: McGraw-Hill, 1969.

[Truesdell, 1980] C. Truesdell. *The tragicomical history of thermodynamics 1822–1854.* New York: Springer, 1980.

[Uffink, 1995] J. Uffink. Can the maximum entropy principle be regarded as a consistency requirement? *Studies in History and Philosophy of Modern Physics, 26*, 223–261, 1995.

[Uffink, 1996] J. Uffink. The constraint rule of the maximum entropy principle. *Studies in History and Philosophy of Modern Physics, 27*, 47–79, 1996.

[Uffink, 1996b] J. Uffink. Nought but molecules in motion. *Studies In History and Philosophy of Modern Physics, 27*, 373–387, 1996.

[Uffink, 2001] J. Uffink. Bluff your way in the second law of thermodynamics. *Studies in History and Philosophy of Modern Physics, 32*, 305–394, 2001.

[Uffink, 2003] J. Uffink. Irreversibility and the second law of thermodynamics. In [Greven et al., 2003, pp. 121–146], 2003.

[Uffink, 2004] J. Uffink. Boltzmann's Work in Statistical Physics. In E.N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2004 Edition), http://plato.stanford.edu/archives/win2004/entries/statphys-Boltzmann.

[Uffink, 2005] J. Uffink. Rereading Ludwig Boltzmann. In P. Hájek, L. Valdés-Villanueva & D. Westerståhl (Eds.) *Logic, Methodology and Philosophy of Science. Proceedings of the twelfth international congress.* London: King's College Publications, pp. 537–555, 2005.

[Uffink, 2006] J. Uffink. Insuperable difficulties: Einstein's statistical road to molecular physics. *Studies in History and Philosophy of Modern Physics*, 2006.

[Uffink and van Lith, 1999] J. Uffink and J. van Lith. Thermodynamic uncertainty relations. *Foundations of Physics, 29*, 655–692, 1999.

[Uhlenbeck and Ford, 1963] G. E. Uhlenbeck. and G. W. Ford. *Lectures in statistical mechanics.* Providence, Rhode Island: American Mathematical Society, 1963.

[Vranas, 1998] P. B. M. Vranas. Epsilon-ergodicity and the success of equilibrium statistical mechanics. *Philosophy of Science, 65*, 688–708, 1998.

[Wald, 2001] R. M. Wald. The thermodynamics of black holes. *Living Reviews in Relativity, 4*, 6. http://www.livingreviews.org/lrr-2001-6.

[Winsberg, 2004] E. Winsberg. Laws and statistical mechanics. *Philosophy of Science, 71*, 707–718, 2004.

[Yourgrau *et al.*,] abc W. Yourgrau, A. van der Merwe, and G. Raw. *Treatise on irreversible thermophysics*, New York: Macmillan, 1966.

[Zermelo, 1896a] E. Zermelo. Ueber einen Satz der Dynamik und die mechanische Wärmetheorie. *Annalen der Physik, 57*, 485–494, 1896. English translation in [Brush, 2003, pp. 382–391].

[Zermelo, 1896b] E. Zermelo. Ueber mechanische Erklärungen irreversibler Vorgänge. *Annalen der Physik, 59*, 793–801, 1896. English translation in [Brush, 2003, pp. 403–411].

QUANTUM STATISTICAL PHYSICS

Gérard G. Emch

1 INTRODUCTION

In search for the headwaters of the Missouri during their 1804–1806 expedition, Lewis and Clark decreed that the river begins at the confluence of three streams — the Jefferson, Gallatin, and Madison rivers — and ends as a main tributary to the mighty Mississippi.

Similarly, and with some of the same arbitrariness, three major headings can be used to mark the beginnings of quantum statistical physics (QSP): Planck's "quantum hypothesis" following his 1900 papers [Planck, 1900a; Planck, 1900b], Gibbs' 1902 book on "statistical mechanics" [Gibbs, 1902], and what is now known as Einstein's 1905 "Brownian motion" [Einstein, 1905b]. Pushing the metaphor into our own days, the power of QSP is manifest in the landscape of condensed matter physics (from solid state physics to astrophysics). The navigation there, albeit often tentative, has brought to shore predictions that have been confirmed with a precision impressive enough to clamor for a consistent explanation. The purpose of this chapter is to point to directions along which such explanations may be found. I begin this search by tracing briefly the course of the three tributaries mentioned above, thus by identifying the initial motivations for QSP.

Planck's long-lasting hesitations indicate how much in advance he was, not only of his own time, but perhaps even of himself; e.g., at first, he had put forward his black-body radiation law on account of the nature of the *body* — little oscillators in the walls — rather than on account of the nature of the *radiation.* As Planck was transposing to the description of electromagnetic waves the counting arguments Boltzmann used in the thermal physics of material bodies, he initially left open the question of whether this was a mere formal analogy, or whether it was one that could be justified from putative interactions between radiation and matter; or whether, yet, this speculative analogy had deeper roots. Planck's reluctance still shows through in the recommendation he wrote in 1913 to support young Einstein's early election to the Prussian Academy of Sciences: "That he may sometimes have missed the target of his speculations, as for example in his hypothesis of the light quanta, cannot really be held against him." While this may be seen as a barb directed to [Einstein, 1905a], note nevertheless that Planck's reference to a quantum *hypothesis* is not a passing accident: he was meticulous in his use of words; consider, for instance the use of "theory", "theorem", and "hypothesis"

Handbook of the Philosophy of Science. Philosophy of Physics
Volume editors: Jeremy Butterfield and John Earman
General editors: Dov M. Gabbay, Paul Thagard and John Woods
© 2007 Elsevier B.V. All rights reserved.

in the title of his 1911 address to the German Chemical Society [Planck, 1911]. Soon thereafter, the rest of the world overcame his scruples: the Nobel prize was awarded to Planck in 1918 for “his discovery of energy quanta”; and to Einstein in 1921 for “his discovery of the law of the photoelectric effect.” For each of them, the *laudatio* calls attention to their respective contributions to the nascent QSP, specifically: the black body radiation for Planck and the specific heat of solids for Einstein; see subsections 2.1 and 2.3 below.

Gibbs’ book [Gibbs, 1902] focuses on *classical* statistical physics. While the basic concepts had been apprehended differently by the German Clausius, the Austrian Boltzmann and the British Maxwell, the American Gibbs proposes that the field has reached the modicum of maturity necessary for a consolidation of the foundations; for axiomatization in other fields, compare with Hilbert [Hilbert, 1900; Hilbert, 1899; Hilbert, 1918], and Einstein [Einstein, 1921]. Even in the classical context, Gibbs’ reluctance to invoke Boltzmann’s ergodic postulate points to the persistence of unresolved issues regarding what Gibbs calls in the very title of his book “the rational foundation of thermodynamics”; for a brief presentation of those aspects of Gibbs’ work that may be most relevant to my purpose here, see [Uffink, 2006, section 5]. It pertains to the present chapter to examine how much of this dichotomy persists in the quantum realm, and the extent to which whatever persists is relevant to the explanatory purposes of QSP.

Einstein’s papers on Brownian motion still reside conceptually in the realm of classical physics. In spite of the neglect in which many mathematicians still held the foundations of probabilistic theories around the turn of the twentieth century (cf. e.g. [Hilbert, 1900, Problem 6]), Einstein’s approach stands as a witness to the fact that stochastic arguments — i.e. arguments involving random processes — had gained currency in the physicists’ marketplace. Einstein’s conclusions were widely (if not universally) accepted at face value as empirical proof of the existence of molecules, as not just computationally convenient small entities or units, but as objects with definite dimensions [Einstein, 1906b]. Furthermore, Einstein’s papers were not the isolated manifestation of a singular genius that the cumbersome title of his first paper might suggest [Einstein, 1905b]. On the one hand, from the physicist’s perspective, it must be noted that Einstein begins his second paper in the sequence with an ackowledgment that he had ignored the earlier contributions of Siedentopf and Gouy who had interpreted the “so-called Brownian motion” [Einstein *dixit*] as caused by the irregular thermal motions of the molecules [Einstein, 1906c; Gouy, 1888]. On the other hand, the modern mathematician will recognize, with the hindsight of practitioners such as Kac and Chandrasekhar, that Smolukowski simultaneously distilled from the same empirical sources the mathematical intuition allowing him to post a claim on what was to become the theory of stochastic processes [Smolukowski, 1906a; Smolukowski, 1916]. Yet, it was only in 1933 that Kolmogorov made precise the essentials of the underlying syntax, namely the mathematical theory of probability [Kolmogorov, 1933]. Even so, an unresolved issue remains to this day as to the proper semantics: von Mises’ collectives [von Mises, 1928] or de Finetti’s subjective assignments [de Finetti, 1937]. I post my stakes — see subsection 3.1 — on the latter issue when considering the extension of the theory of probability to the quantum realm, with special regard to the specific demands of QSP.

As this essay opens, the question arises as to whether the confluence of three streams of interest compounds the foundational problems of each of them or, on the contrary, whether they can be brought to inform one another. I aim my argument towards the latter view, although I am not oblivious to such ubiquitous problems as questioning what elements of reality should — or should not — be ascribed to individual microscopic quantum systems. As part of the larger problem of the reduction of thermodynamics by statistical mechanics, I consider specifically the question whether and how QSP can claim to explain the collective properties of many-body systems: it does postulate a quantum description at the microscopic level, while it has not obtained as yet an ontological grasp of the individual components of these systems. In my presentation I follow Einstein's admonition: "If you want to find out anything from the theoretical physicists about the methods they use ... don't listen to their words, fix your attention on their deeds." [Einstein, 1933].

2 EARLY SUCCESSES

In [Jammer, 1966] Max Jammer provides much of the specific historical documentation pertaining to the beginnings of quantum theory; and he discusses some of the ensuing debates in [Jammer, 1974]. Here, I start with a discussion of the early pragmatic successes of QSP, with special attention to two aspects: their classical mooring in the high temperature regime; and the understanding QSP gives of the particle-wave duality. Both of these aspects illustrate the added insight gained from the contextual differences coloring the answers to the same questions when asked in QSP rather than in the quantum theories of, say, the Bohr atom or scattering processes; compare with Mara Beller's perspective on the making of the quantum revolution [Beller, 1999].

2.1 Planck's interpolating formula for black-body radiation

The *experimental evidence* available to Planck was the spectral density $\rho_T(\nu)$ of the energy per unit volume of electromagnetic radiation, as a function of its frequency ν, when electromagnetic radiation is in equilibrium with a black-body at temperature T. In [Planck, 1900a; Planck, 1900b], Planck proposes to fit these data with the formula

$$(1) \quad \rho_T(\nu) = A\,\frac{h\nu}{e^{\frac{h\nu}{kT}}-1} \quad \text{with} \quad A = \frac{8\pi\nu^2}{c^3}$$

where c is the speed of light, $k = R/N_{Av}$ is the Boltzmann constant, R is the universal gas constant and N_{Av} is the Avogadro number. In addition, *a new* *constant* enters the formula, h, known nowadays as the Planck constant. While Planck himself would pretend that (1) had been a "lucky guess" such a formula could not have come into existence in a conceptual vacuum.

Two qualitative laws had been identified by Wien [1894], Stefan [1879] and Boltzmann [1884]. The *Wien displacement law* states that

$$(2) \quad \rho_T(\nu) = \nu^3 \, f(\frac{\nu}{T})$$

where f is some undetermined function, satisfying the condition that the following integral converges

$$(3) \quad \frac{1}{V} E(T) = \int_o^\infty d\nu \rho_T(\nu)$$

which expresses the density, per unit volume, of the energy of the radiation at temperature T. Upon inserting (2) in (3), one receives immediately the *Stephan–Boltzmann* law:

$$(4) \quad E(T) = \sigma T^4$$

where σ is a constant. Planck's proposal complies with these laws.

Two analytic expressions (or "laws") had been proposed, which specify the function f in (2). One law, due to Wien [1896], reads:

$$(5) \quad \rho_T(\nu) = \alpha \, \nu^3 \, \exp^{-\gamma \, \frac{\nu}{T}} \quad .$$

With α and γ being two constants, this law had been confirmed empirically in the range where ν/T is *large.* In contrast, the other law, due to Rayleigh [1900], see also Jeans [1905a], is:

$$(6) \quad \rho_T(\nu) = \frac{8\pi}{c^3} \, \nu^2 \, kT$$

which had been confirmed empirically in the range where ν/T is *small.*

Clearly, (1) interpolates analytically between the Wien and Rayleigh–Jeans formulas; and it gives a quantitative meaning to the conditions that ν/T be "large" (resp. "small"), namely $\nu/T \gg k/h$ (resp. $\nu/T \ll k/h$). In the intermediate range, Planck's interpolating formula fits experimental results very well, both qualitatively and quantitatively.

Planck's colleagues could not fail to be impressed and Planck's triumph would have been complete had he been able to explain his formula from first principles, at least to the considerable extent with which (2) to (6) could be understood. Instead, Planck has to resort to "an act of desperation" — his own words [Jammer, 1966] — and he constructs, after several attempts, a heuristic model in which the radiation exchanges energy in discrete *quanta* with putative "resonators" in thermodynamical equilibrium within the walls. The model suffers from several shortcomings — among them Planck's adaptation of Boltzmann's counting — and much uncertainty concerning its theoretical status:

> Walter Nernst ... initially disliked quantum theory, claiming that it was 'really nothing else than an interpolation formula ... only a rule for calculations ... but has proven so fruitful by the work of Planck ... and ... of Einstein that the duty of science is to take it seriously and to subject it to careful investigations'. [Jammer, 1966, p. 59]

The consensus that later settled in the physics community is that *any* attempt — Planck's included — to derive (1) from first principles would be doomed to failure: (1) *is* a fundamental or primary law, i.e. one that is not to be explained, but the consequences of which ought to be explored.

2.2 Einstein's fluctuation formula and the particle-wave duality

For a start, Einstein notes two shortcomings in Planck's derivation. The first is formal, but nevertheless essential: Planck's account does not conform to Boltzmann's statistical counting as closely as Planck suggests. The second is pointed out in [Einstein, 1906a]: Planck's treatment involves an inconsistency between: (a) his use of the (classical) Maxwell theory of electromagnetism to compute the average energy of a resonator in a radiation field; and (b) the assumption that the energy of a resonator can change only discontinuously. Together with other empirical problems — among which the photoelectric effect [Einstein, 1905a] — these difficulties led Einstein to propose that, while Planck's radiation formula (1) has incontestable empirical merits, the "quantization" itself is to be looked for in the radiation field rather than in a dubious mechanism of interaction with the walls. Einstein's criticism raises, in the same volley, the question of whether light is wave-like as accounted for by Maxwell's electromagnetic theory; or whether it is particle-like as Newton's theory had it before its purported falsification in interference experiments conducted in the early nineteenth century .

Einstein's fluctuation formula [Einstein, 1909a] proposes that light should be viewed *simultaneously* as *both* particle *and* wave; specifically:

SCHOLIUM 1. Let the Planck's spectral density $\rho_T(\nu)$ in (1) be interpreted as the average energy $\langle u_T(\nu)\rangle$ of quantum oscillators of frequency ν of the radiation in thermal equilibrium at temperature T . Then for all values of $h\nu/kT$, the energy fluctuation $\langle(\Delta u)^2\rangle = kT^2\partial_T\langle u_T(\nu)\rangle$ is the sum of two terms

$$\langle(\Delta u)^2\rangle = \langle(\Delta u)^2\rangle_p + \langle(\Delta u)^2\rangle_w \quad \text{where}$$

$$(7) \qquad \left\{ \begin{array}{l} \langle(\Delta u)^2\rangle_p = \langle u_T(\nu)\rangle\, h\nu \\ \\ \langle(\Delta u)^2\rangle_w = \langle u_T(\nu)\rangle^2\, \frac{c^3}{8\pi\nu^2} \end{array} \right\} \quad \text{and} \quad \langle(\Delta u)^2\rangle_p \,/\, \langle(\Delta u)^2\rangle_w = \exp^{\frac{h\nu}{kT}} - 1$$

Hence, the particle-like contribution $\langle(\Delta u)^2\rangle_p$ dominates when $h\nu/kT >> 1$, and the wave-like contribution $\langle(\Delta u)^2\rangle_w$ dominates when $h\nu/kT << 1$. In this interpretation, *the particle-wave duality is thus a matter of degree,* rather than an alternative between the two mutually exclusive horns of a dilemma.

Less of a conceptual problem in QSP, this duality becomes more difficult to master in other empirical contexts where one may prefer to view a photon *either* as a particle *or* as a wave packet. Moreover, this duality has since been extended to all (sub-atomic) particles; e.g. phenomena usually associated with waves, such as diffraction of beams of light, have been observed as well with beams of electrons and then neutrons; cf. e.g. [Jammer, 1966, pp. 249-253]; or for an update [Rauch, 2005]. In other circumstances, one prefers to use a particle language, as for instance in the description of the photo-electric effect [Einstein, 1905a]; as reported in most QM textbooks, a photon impinging on a metallic surface causes an electron to be expelled; or in atomic spectroscopy, a particle — the atom — emits a beam of light; cloud- and bubble-chambers have since let us visualize interparticle collisions; and yet their description in scattering theory uses the so-called wave operator; cf. e.g. [Amrein *et al.*, 1977]. In the light of this duality, and following upon the speculations of Einstein and de Broglie, physicists have learned to adapt their language to the aspect they wish to emphasize. Yet, the persistent arguments about "self-interference" show that some residual ambiguities have yet to be resolved; cf. the long debate extending from [Taylor, 1909] to [Aichele *et al.*, 2005], and surely beyond.

Upon returning to the early manifestations of QSP, one ought to mention that the Einstein fluctuation formula (7) above, as well as the explanation of the temperature dependence of the specific heat of solids — see subsection 2.3 below — motivate the Ehrenfests' suggestion [Ehrenfest and Ehrenfest, 1911] that in statistical mechanics, quantum behaviour manifests itself mostly at low temperatures, whereas classical behaviour emerges at high temperatures. The fact is that in many expressions, such as the Planck distribution (1), the Planck constant h and the temperature T appear together in a factor h/T, or in the form used in the sequel, $\hbar\beta$; hence in these expressions the "classical limit" $h \to 0$ and the "high-temperature limit" $T \to \infty$ are included in $(\hbar\beta) \to 0$. All refer to cases where the relevant energies, or energy densities, are extremely large when measured in the scale determined by the numerical value of the Planck constant.

2.3 Debye's specific heat of solids below the classical regime

For the purpose of this subsection, the situation down in the field is that Dulong & Petit (1819) had proposed an argument to the effect that the specific heat — measured in calories per mole per degree — ought to be the same for all solids: $3R$ where R is the universal gas constant. Yet, it later became apparent that this "constant" could decrease dramatically with temperature, so much so that by the end of the nineteenth century, the experimental data led to the *conjecture* that the specific heat of solids becomes vanishingly small as the temperature approaches absolute $0\,K$. In the meantime, the discovery of X-rays by Roentgen (1895) had allowed several experimentalists — Ewald (1911), and at the suggestion of von

Laue, Friedrich and Knipping (1912) — to obtain diffraction patterns corroborating speculations that crystalline solids are regular lattices, at the vertices of which sit the atoms.

As no classical explanation of the observed drastic temperature dependence of the specific heat seemed forthcoming, Einstein and Debye offered the following model; cf. [Einstein, 1907; Einstein, 1911b; Debye, 1912].

The starting point is (1) above, the Planck formula for black-body radiation, now reinterpreted in terms of the vibrational modes of a solid at temperature T :

(8) $$U(T) = \int d\nu\, g(\nu)\, U(\nu,T) \quad \text{with} \quad U(\nu,T) = \frac{h\nu}{e^{\frac{h\nu}{kT}} - 1} \quad \text{and} \quad \int_0^\infty d\nu\, g(\nu) = 3N$$

where N is the number of 3-dimensional oscillators in the solid. Where Einstein had assumed that g is concentrated on a fixed frequency ν_o, Debye chooses for g the simplest vibrational distribution that takes into account that in a crystal, say of volume V, the vibrations have a minimal wavelength of the order of the interatomic distance in the lattice:

(9) $$g(\nu) = G\left\{ \begin{array}{lll} 1 & \text{if} & 0 \leq \nu \leq \nu_o \\ 0 & \text{if} & \nu > \nu_o \end{array} \right\} \quad \text{with} \quad G = \frac{12\pi\nu^2}{s^3} V \quad .$$

G takes into account that vibrations are now sound waves rather than electromagnetic waves — compare with $A = \frac{8\pi\nu^2}{c^3}$ in (1) — thus s is now the speed of sound, instead of the speed c of light; and the replacement of $8\pi = 2{\cdot}4\pi$ by $12\pi = (2{+}1){\cdot}4\pi$ reflects the fact that sound-waves in solids have, in addition to the two transverse polarizations also present in light, a third degree of freedom, namely longitudinal modes. These hypotheses entail the following consequence.

SCHOLIUM 2. There exists a temperature Θ, such that the specific heat satisfies

(10) $$C_V \simeq \left\{ \begin{array}{ll} 3R & \text{for } T \gg \Theta \\ \frac{12}{5}\pi^4 R(\frac{T}{\Theta})^3 & \text{for } T \ll \Theta \end{array} \right. .$$

Hence, the Debye model differentiates between two regimes: at high temperatures it recovers the Dulong–Petit law; and it predicts that as the temperature approaches $0\,K$, the specific heat vanishes according to $C_V \sim T^3$. In this model, the temperature Θ, now called the *Debye temperature,* depends on the solid considered through the cut-off frequency ν_o, and thus on the speed of sound in that solid and on its density N/V. The numerical value of Θ gives a quantitative estimate — for details, see subsection 6.1 — of what is meant by high and low temperature regimes for the specific heat of crystalline solids. Moreover, in Debye's model, C_V decreases monotonically and continuously over the whole range of temperatures $T \in \mathbb{R}^+$.

As a last comment on the passage from (1) to (8), note that by analogy with the *photons* as the quanta of light, the elementary sound vibrations in solids are viewed as quanta, now known as *phonons*.

2.4 BE-condensation: the long haul

When taking seriously the idea that the microscopic picture of the macroscopic world may be a quantum one, the most immediate question is to obtain the corresponding description of a quantum ideal gas; this came to be known as the Bose—Einstein gas, or simply the Bose gas [Bose, 1924; Einstein, 1924]. The starting point is the grand canonical partition function $Z(\Lambda, T, \mu)$ of an assembly of identical massive particles of mass m in equilibrium at temperature T and chemical potential μ; this assembly is enclosed in a cubical box of volume $\Lambda \subset \mathbb{R}^3$, with periodic boundary conditioms. As these particles are non-interacting, the total energy is the sum of their individual energies $\epsilon_k = \hbar^2 |k|^2/2m$, where $k \in Z^3$. The quantum hypothesis is that the Planck distribution (1) applies here so as to entail (with $\beta = 1/kT$) :

$$(11) \quad Z(\Lambda, T, \mu) = \prod_{k \in Z^3} (1 - \exp^{-\beta(\epsilon_k - \mu)})^{-1} \quad .$$

From this formula, one computes the specific volume v and the pressure P, according to the rules learned in classical statistical mechanics; the so-called activity is defined as $z = \exp(\beta\mu)$:

$$(12) \quad v^{-1} = z\partial_z \frac{1}{|\Lambda|} \ln Z(\Lambda, T, \mu) \quad \text{and} \quad \beta P = \frac{1}{|\Lambda|} \ln Z(\Lambda, T, \mu) \quad .$$

The problem is thus stated completely, although the consequences of (11–2.12) are not easy objects to compute directly. The solution involves a mathematical excursion through some classical analysis, and the reward is a nice physical bounty: a phase transition with the onset of a condensed phase at very low temperatures; not your classical ideal gas!

The necessary classical analysis — now widely available, cf. e.g. [Whittaker and Watson, 1927, p.280, ex. 7], [Erdélyi, 1953, I, pp. 27–30], or for some historical perspective [Truesdell, 1945] — was already known to our pioneers, and they did recognize that in the limit $\Lambda \uparrow \mathbb{R}^3$, these sums reduce to:

$$(13) \quad \left.\begin{aligned} v^{-1} &= 4\pi \int_o^\infty dp\, p^2\, z[\exp(\hbar^2 p^2/2mkT) - z]^{-1} \\ \beta P &= 4\pi \int_o^\infty dp\, p^2\, \ln[1 - z\, \exp(-\hbar^2 p^2/2mkT)] \end{aligned}\right\}$$

which are known in classical analysis as Appell integrals, namely

$$(14) \quad \left.\begin{aligned} v^{-1} &= \lambda^{-3} g(\tfrac{3}{2}, z) \\ \beta P &= \lambda^{-3} g(\tfrac{5}{2}, z) \end{aligned}\right\} \quad \text{with} \quad \left\{\begin{aligned} \lambda^2 &= 2\pi\hbar^2/mkT \\ g(s,z) &= \tfrac{z}{\Gamma(s)} \int_o^\infty dt\, \tfrac{t^{s-1}}{\exp^t - z} \end{aligned}\right. \quad .$$

For every s with $Re(s) > 0$, g defines a function of z which is analytic in the cut complex plane $C\backslash[1, \infty)$. For $|z| < 1$ and $Re(s) > 0$, one receives the well-studied Lerch zeta functions which can be expanded in power series

$$(15)\quad g(s,z) = z\,\zeta(s,z) \quad \text{with} \quad \zeta(s,z) = \sum_{n=0}^{\infty} z^n\,(n+1)^{-s} \quad .$$

For $z = 1$ and $Re(s) > 1$ the above series converges to the Riemann zeta function $\zeta(s)$. Note that the values $s = \frac{3}{2}$ and $s = \frac{5}{2}$ — which are needed in (14) — fall within this range. Moreover $g(\frac{3}{2},\cdot) : z \in (0,1) \mapsto R^+$ is smooth, strictly increasing, with $\lim_{z\to 1} g(\frac{3}{2}, z) = \zeta(\frac{3}{2}) = 2.612\ldots$. The problem is thus mathematically under complete control.

Now to the physics. This divides into two steps.

The *first* step is easy: it considers the high temperature and low density regime, where $\lambda^3\,v^{-1} < g(\frac{3}{2},1) = \zeta(\frac{3}{2})$. In particular, by straightforward 1st-order power expansion:

$$(16)\quad \text{for } \lambda^3\,v^{-1} << 1 \;:\; P\,v = kT\,[1 - 2^{-5/2}(\lambda^3\,v^{-1}) + \ldots\,] \quad .$$

Hence, in this high temperature and low density regime, the quantum gas behaves asymptotically like the classical ideal gas of Boyle/Mariotte/Gay–Lussac. This is yet another confirmation of the Ehrenfests' remark according to which the classical limit obtains in QSP as a high temperature limit; note indeed that the so-called *thermal wavelength* λ that appears in (14) satisfies $\lambda \sim \hbar\,\beta^{\frac{1}{2}}$, i.e. in this problem again, the limits $T \to \infty\,(\Leftrightarrow\; \beta \to 0)$ and $\hbar \to 0$ have formally the same effect.

The *second* step in the treatment of the problem is where the bounty is to be found. The question is how to go beyond the above regime, i.e. beyond the unnatural limit

$$(17)\quad \lambda^3 v^{-1} = \zeta(\frac{3}{2}) \quad ,$$

a restriction no actual gas should be expected to respect. Mathematically, this limiting condition seems to appear as the consequence of the breakdown of analyticity in (14) that begins at $z = 1$. Physically, the problem appears because the limit $|\Lambda| \to \infty$ has been taken too carelessly.

Let us therefore return to the expression of v^{-1} when $|\Lambda| < \infty$. We have then, with $\langle n_k\rangle$ denoting the average number of particles in mode k :

$$(18)\quad \frac{1}{|\Lambda|}\sum_{k\in Z^3}\langle n_k\rangle = \frac{1}{|\Lambda|}\sum_{k\in Z^3, k\neq 0}\langle n_k\rangle + \frac{1}{|\Lambda|}\,\frac{z}{1-z} \quad .$$

As the $\langle n_k\rangle$ with $k \neq 0$ are well-behaved as $z \to 1$, the separation of (18) into two terms suggests that we take simultaneously the limits $|\Lambda| \to \infty$ and $z \to 1$ in such a manner that the second term in (18) approaches a finite limit, say ${v_o}^{-1}$, resulting in the replacement of (14) by:

$$(19)\quad \left.\begin{aligned} v^{-1} &= \lambda^{-3}\,\zeta(\tfrac{3}{2}) + {v_o}^{-1} \\ \beta\,P &= \lambda^{-3}\,\zeta(\tfrac{5}{2}) \end{aligned}\right\} \quad .$$

The above limiting procedure, interpreted as

$$(20)\quad {v_o}^{-1} = \lim_{|\Lambda|} \frac{1}{|\Lambda|} \langle n_o \rangle \quad ,$$

leads to a *macroscopic occupation of the ground state* $k = 0$; the theory does not predict the value of v_o : it may depend on the temperature. Note that the pressure P in (19) depends on temperature only (namely through λ). The state of the system described by (19) is called its *condensed phase*; the transition to this phase from the normal phase $\lambda^3 v^{-1} < \zeta(\frac{3}{2})$ is referred to as the *Bose–Einstein condensation,* (or *BEC*) and its appearance at low temperature is a prediction of purely quantum origin, one that has no equivalent in the classical world.

This begs for an instantiation in the world of the laboratory. At low temperature a superfluid phase appears in ^{4}He. The density at the onset of this phenomenon is about $\rho \simeq .178$ g/cm^3. Upon taking into account the value of the Avogadro number, one receives $v^{-1} \simeq 2.7 \cdot 10^{23}$ cm^{-3}, from which (17) gives a thermal wavelength $\lambda \simeq 4.6 \cdot 10^{-8}$ cm which is not unreasonable for a quantity that is to be interpretated as a measure of the interparticle distance. To this value corresponds, via the definition of λ in (14), a temperature $T \simeq 3.2\,K$. The experimental value of the temperature at the onset of the superfluid phase in ^{4}He is $T \simeq 2.2\,K$, a rather remarkable fit, considering how crude the model is. Moreover, the thermodynamics of the model can be worked out — cf. e.g. [Huang, 1965] — and shows that the specific heat $C_v(T)$ at first increases monotonically from $C_v(0) = 0$ to exceed the classical value 3/2 but then experiences a sharp peak — a discontinuity in the first derivative — from which it decreases monotonically to $\lim_{T\to\infty} C_v(T) = 3/2$. The specific heat of ^{4}He also exhibits such a singularity, albeit more pronounced: it is logarithmic; hence its name $\lambda -$ *point*, as the graph of the specific heat as a function of temperature looks like the Greek lower case letter lambda.

All this represented a great success in the the mid-1920s. The next batch of problems appeared when the theory tried to account for the fact that ^{4}He is not a gas, but a liquid; for this, the *ideal* gas assumption of the model is quite unrealistic: a liquid is not made of non-interacting particles. Putting the interactions into the theory proved to be a formidable problem, long compounded by the experimental fact that ^{4}He was the only substance recognized to exhibit Bose–Einstein condensation: theoreticians had no variable parameter to guide and adjust their speculations. Following up on a proposal made in the late 1950s, the situation changed drastically during the 1980s and 1990s with the advent of micro-Kelvin technology which allowed BEC to be observed in atomic gases in harmonic traps; for two deep, but very different, reviews, cf. [Lieb, 2001] and [Pitaevskii and Stringari, 2003]; and for a brief overview [Emch and Liu, 2002, subsection 14.2.2].

The account in this subsection was limited mostly to the macroscopic, thermodynamical aspects of BEC in its infancy; in subsection 5.2 below, a C$^*-$algebraic treatment of the Bose–Einstein model is discussed in connection with the appearance in QSP of the modular structures to be associated to the equilibrium KMS condition.

2.5 Beyond the Bohr atom: the Thomas–Fermi model

The entry in quantum mechanics of Schrödinger wave-mechanics (1926) was marked by a resounding success: the physics community could recognize immediately the application of a then already well-established method to a new realm; the theoretical explanation of the energy spectrum of the hydrogen atom was reduced to solving an eigenvalue problem in a differential equation. Every entry-level text in quantum mechanics presents this derivation.

And yet, beyond the Bohr atom, the solution of the Schrödinger equation for an atom with even a few electrons turned to be an insurmontable task: the electrons are charged particles and while the interaction between a single electron and the nucleus had been rigorously accounted for in the hydrogen atom, one could not deal analytically with the mutual electromagnetic interactions between the electrons.

Very soon thereafter, Thomas [1927] and Fermi [1927] came up with a semi-classical model in which two ingredients enter. The first is the ground state electron density ρ which is assumed to be spherically symmetric and normalized by the condition

$$(21)\quad 4\pi \int_o^\infty dr r^2 \rho(r) = Z$$

where eZ is the charge of the nucleus. The second is the average electric potential $\Phi(r)$ in the atom. These two ingredients are assumed to satisfy the *classical* equation, the Poisson equation of electrostatics

$$(22)\quad \Delta\Phi \equiv \frac{1}{r}\frac{d^2}{dr^2}(r\Phi) = 4\pi e\rho \quad \text{with} \quad \lim_{r\to 0}\Phi(r) = eZ \quad .$$

And yet the model has a quantum aspect to account for the Pauli exclusion principle; this is the so-called Fermi–Dirac statistics that had been proposed just the previous year [Fermi, 1926]. Here, this shows up in:

$$(23)\quad n(r,p) = \begin{cases} 2h^{-3} & \text{if} \quad \epsilon := \frac{1}{2m} - e\Phi < \epsilon_o \\ 0 & \text{if} \quad \epsilon > \epsilon_o \end{cases} \quad .$$

from which one gets, by integration over p (upon putting $\epsilon_o = 0$), that ρ satisfies

$$(24)\quad \rho(r) = \begin{cases} \frac{8\pi}{3h^3}(2me\Phi)^{3/2} & \text{if} \quad \Phi > 0 \\ 0 & \text{if} \quad \Phi < 0 \end{cases} \quad .$$

Clearly, the model is conceptually inconsistent, with stakes in each of the classical and the quantum realms. Yet, in my student days this model was a routine staple of the quantum mechanics curriculum [Schiff, 1955; Landau and Lifshitz, 1958a; Messiah, 1960] as it can be solved without any further assumptions than those listed above; the solution is exact up to the fact that it requires a numerical computation well within the realm of a controllable approximation.

Upon using the numerical values of the Planck constant h, the charge e and the mass m of the electron, the model predicts that the radius of the atom, taken to be the radius of the sphere that contains all the electrons but one increases monotonically from $2.2 \cdot 10^{-8}$ cm for $Z = 25$, to $2.8 \cdot 10^{-8}$ cm for $Z = 100$. The order of magnitude is correct. This can be counted therefore as an early success of quantum theory.

However, one should expect that such a crude model does not tell the whole story. Indeed: (1) the predicted increase stops at $Z = 55$ (corresponding to the cesium atom) after which the radius decreases, albeit slowly; (2) when looked at more closely, the model yields an electron density that has unreasonable properties both very close and very far from the nucleus. Besides, the model needs serious reconsiderations to explain the existence of stable molecules or to accommodate a relativistic treatment. These problems never completely left the scene of theoretical physics, but remained somewhat in the background for about half-a-century, until rigorous analytic methods clarified the sense in which the model is asymptotically exact and may be used to study the stability of atoms, molecules and even stars; cf. [Lieb and Simon, 1977; Lieb, 1982a; Lieb, 1990]; see also [Catto *et al.*, 1998; Le Bris and Lions, 2005].

2.6 White dwarfs: the Chandrasekhar bound

Returning to the quantum ideal gas discussed in subsection 2.4, let us examine now the Fermi gas. Instead of (11), start with the partition function

$$(25)\quad Z(\Lambda, T, \mu) = \prod_{k \in Z^3} (1 + \exp^{-\beta(\epsilon_k - \mu)})$$

which now entails in the limit $\Lambda \uparrow Z^3$, instead of (13):

$$(26)\quad \left.\begin{aligned} v^{-1} &= 4\pi \int_o^\infty dp\, p^2\, z[\exp(\hbar^2 p^2/2mkT) + z]^{-1} \\ \beta\, P &= 4\pi \int_o^\infty dp\, p^2\, \ln[1 + z\, \exp(-\hbar^2 p^2/2mkT)] \end{aligned}\right\} \quad .$$

In the high temperature and low density regime — $\lambda^3 v^{-1} \ll 1$ — one recovers again an asymptotic expansion, the leading term of which is the classical ideal gas:

$$(27)\quad for\ \lambda^3\, v^{-1} << 1\ :\ P\, v \simeq kT\, [1 + 2^{-5/2}(\lambda^3\, v^{-1}) + \ldots\,] \quad .$$

Again, up to the sign of the correction, this is very similar to the Bose–Einstein result (16): it also coincides asymptotically with the classical ideal gas as T becomes large.

In the low temperature and high density regime — $\lambda^3 v^{-1} \gg 1$ — the situation differs drastically from what it was in subsection 2.4: whereas bosons tend to congregate, no two fermions are allowed in the same state on account of the Pauli exclusion principle. Recall that in chemistry, this is the principle that underpins a quantum explanation for the Mendeleev table of elements. In QSP the Pauli principle is visible through (26): in the ground state of the system, the fermions occupy the lowest possible energy states up to a finite energy, called the Fermi–energy

$$(28)\quad \epsilon_F = \frac{\hbar^2}{2m}\left[(3\pi^2)\, v^{-1}\right]^{\frac{2}{3}} \quad .$$

For temperatures such that $kT \ll \epsilon_F$ the momentum distribution will be

$$(29)\quad < n_p >= \begin{cases} 1 & \text{for} \quad (|p|^2/2m) \lesssim \epsilon_F \\ 0 & \text{for} \quad (|p|^2/2m) \gtrsim \epsilon_F \end{cases}$$

with a steep sigmoid of narrow breadth kT around ϵ_F. This regime is called the *degenerate Fermi gas.* To characterize this regime, rewrite $kT \ll \epsilon_F$, with ϵ_F as in (28), as:

$$(30)\quad \beta\, v^{-\frac{2}{3}} \gg \left[\frac{\hbar^2}{2m}(3\pi^2)^{\frac{2}{3}}\right]^{-1}$$

which gives a quantitative meaning to the expression *low temperature and high density* regime; for instance, this yields a useful first approximation for the gas of electrons in metals at usual temperatures. The condition $kT \ll \epsilon_F$ corresponds to $\lambda^3 v^{-1} \gg 1$ and in this regime (26) entails

$$(31)\quad Pv \simeq \frac{2}{5}\epsilon_F\left[1 + \frac{5\pi^2}{12}(\frac{kT}{\epsilon_F})^2 + \ldots\right] \quad \text{i.e.} \quad \lim_{\frac{kT}{\epsilon_F}\to 0} Pv^{\frac{5}{3}} = \frac{2}{5}(3\pi^2)^{\frac{2}{3}}\frac{\hbar^2}{2m} \quad .$$

Hence, at fixed density, the pressure approaches a strictly positive constant as $T \to 0$, in marked contrast with the behaviour of the classical ideal gas (see (27)) where $T \to 0$ implies $P \to 0$.

Less mundane examples are provided by celestial objects, white dwarfs and neutron stars. With a temperature similar to that of the sun, i.e. $10^7 K$ to $10^8 K$ in the center, and a mass of the same order of magnitude as the sun, the white dwarfs have a very high density, about 10^6 to 10^7 times that of the sun. They are stars where all the hydrogen fuel has been burned, and thus they are constituted of completely ionized helium atoms. From these hypotheses on the composition and condition of a white dwarf, one computes the density of the electron gas, and then from (28) the corresponding Fermi energy ϵ_F which, when expressed in terms of $T_F = \epsilon_F/k$, turns out to give $T_F \simeq 10^{11} K$. Hence $T \ll T_F$ and it is consistent to assume — as R.H. Fowler did already in 1926 [Fowler, 1926] — that the electron assembly in the white dwarfs may be described as a degenerate Fermi gas, and that it is the enormous pressure in such a gas that prevents the star from gravitational collapse. It is however true that at such density and pressure, electrons must be treated relativistically, i.e. $\epsilon = \sqrt{(pc)^2 + (mc^2)^2}$ instead of $\epsilon = p^2/2m$. This brings about all sorts of analytic difficulties, among which is a change from 5/3 towards 4/3 in the power of v in (31). In the course of his computations of this effect, Chandrasekhar [1931a] remarked that since the gravitational pressure is governed by the mass of the star, the latter would collapse if the mass were to become too large; he actually evaluated this critical mass M_{max} to be

(32) $M_{max} \simeq (3\pi)^{\frac{1}{2}}(\frac{\hbar c}{G})^{3/2}(\mu m_N)^{-2} \simeq 1.4\, M_{\odot}$

where (in cgs units) $\hbar = h/2\pi$ with $h \simeq 6.62 \times 10^{-27}$ ergs cm is the Planck constant, $c \simeq 3 \times 10^{10}$ cm/sec is the speed of light, $G \simeq 6.67 \times 10^{-8}$ dyn cm^2 g^{-2} is Newton's gravitational constant, $m_N \simeq 1.66 \times 10^{-24}$ g, μ is the number of nucleons per electron; here $\mu = 2$ since the star is supposed to have used its hydrogen supply, and be made of ${}^4_2$He. Finally, to reduce the result in astronomic units, $M_{\odot} \simeq 1.99 \times 10^{33}$ g is the mass of the sun. Astronomers today refer to the maximum mass M_{max} as the *Chandrasekhar limit* [where mathematicians would speak of a "bound"].

Chandrasekhar's original derivation is mathematically correct, yet somewhat cumbersome. Already by the end of 1932, L.D. Landau [1932] presented a more elementary argument, *and* in addition, upon hearing of the discovery of the neutron, he applied the above formula to then putative neutron stars.

To have included these predictions here among the early "successes" of QSP may be justified only by hindsight. When they appeared in the early 1930s they and their consequences caused quite a wave, on the crest of which rode A.S. Eddington, an astronomer of commanding authority, who spoke of a *reductio ab absurdum* calling for the interposition of an as yet unknown fundamental theory: for him a massive star ($M > M_{max}$) collapsing to a black hole was heresy, and he was in a position not to mince his words about it. Eddington's fierce attack on a junior colleague did not cause Chandrasekhar to recant; unconvinced, Chandra nevertheless decided to turn to other astronomical problems until the late 1950s [Chandrasekhar, 1958] and early 1960s, when his speculations, and Landau's, found observational confirmations.

For the unfolding of the resolution of the Chandrasekhar–Eddington conflict, cf. e.g. [Shapiro and Teukolsky, 1983], the title of which already indicates the complete extent to which Chandrasekhar was ultimately vindicated. A pristine, yet nontechnical presentation of the physics of the Chandrasekhar bound may be read in [Thorne, 1994, chapter 4]; the story of the neutron stars, albeit more involved is also told there in [Thorne, 1994, chapter 5]; for the fundamental technical support, cf. [Weinberg, 1972, chapter 11].

3 AXIOMATIC PRUNINGS

Usually, either one of two reasons prompts the process of axiomatization. The first is the search for the soul — some would say the skeleton — hidden inside the aleatory appearances of the body: a ritual of purification. The second is the need for fundamental changes when a theory faces increasingly insuperable limitations. Both of these reasons motivate the developments I retrace in the present section; as I write this I am reminded of the essential tensions described elsewhere in [Segal, 1990].

It is an interesting coincidence that the early 1930s saw the almost simultaneous — albeit independent — axiomatizations of two of the ingredients of QSP:

Quantum mechanics with von Neumann's treatise [von Neumann, 1932c]; and *Statistics,* a.k.a. probability and stochastic processes, with Kolmogorov's paper [Kolmogorov, 1933]. As both of these belong to other chapters of this Handbook, only a few words will suffice here.

3.1 Kolmogorov's and von Neumann's formalisms compared

In a nutshell, Kolmogorov's syntax for probability starts with a seminal description of measure theory: a triple $\{\Omega, \mathcal{E}, \mu\}$ is given where $\mathcal{E}$ is a $\sigma-$algebra of measurable subsets of a set Ω, and μ is a countably additive function

$$(33)\quad \mu : E \in \mathcal{E} \mapsto \mu(E) \in \mathbb{R}^+ \quad \text{with} \quad \mu(\Omega) = 1$$

i.e. μ is a *probability measure* on $\{\Omega, \mathcal{E}\}$. μ naturally extends to a functional on the algebra $\mathcal{A} = \mathcal{L}^\infty(\Omega, \mathcal{E}, \mu)$ of all essentially bounded functions $A : \Omega \to \mathbb{C}$:

$$(34)\quad \mu : A \in \mathcal{A} \mapsto \mu(A) = \iint_\Omega d\mu(\omega)\, A(\omega) \in \mathbb{C}\ .$$

Hereafter, I will refer to this extension as a *classical state.*

Similarly, von Neumann's syntax involves a triple: $\{\mathcal{H}, \mathcal{P}, \psi\}$ where $\mathcal{P}$ is the orthomodular lattice of all closed subspaces of a Hilbert space $\mathcal{H}$, ψ is a countably additive positive function

$$(35)\quad \left.\begin{array}{ll} \psi : P \in \mathcal{P} \mapsto \psi(P) \in \mathbb{R}^+ & \text{with} \quad \psi(I) = 1 \quad \text{and} \\ \psi(\sum_n P_n) = \sum_n \psi(P_n) & \forall \{P_n\} \subset \mathcal{P} \text{ such that } n \neq m \models P_n \perp P_m \end{array}\right\} .$$

I shall refer to any such funtion ψ as a quantum state. Gleason's theorem asserts in particular — see below for a complete statement — that for every quantum state ψ there exists a density operator, i.e. a positive operator ρ of unit trace, such that ψ extends to the W^*-algebra $\mathcal{B} = \mathcal{B}(\mathcal{H})$ of all bounded linear operators from $\mathcal{H}$ into itself:

$$(36)\quad \psi : B \in \mathcal{B} \mapsto \psi(B) = \mathrm{Tr}\rho\, B \in \mathbb{R}\,.$$

When working within the von Neumann formalism, I will identify any closed subspace $P \subseteq \mathcal{H}$ and the projector $P \in \mathcal{B}(\mathcal{H})$ on this subspace; I will indifferently refer to ψ or to ρ as a *state* on $\mathcal{B}$; and I will refer to the restriction of ψ to $\mathcal{P}$ as a quantum measure. I will also follow the physicist's custom of referring to ρ as a *density matrix,* thus ignoring the mathematician's distinction between an operator and its expression in a specified (orthonormal) basis.

The mathematical similarities and differences between the classical and quantum realms are emphasized by the Koopman formalism of classical mechanics; cf. e.g. [Emch and Liu, 2002, pp. 255, 267]. This formalism — actually a precursor of the GNS construction — associates to $\{\Omega, \mathcal{E}, \mu\}$ the Hilbert space $\mathcal{H} = \mathcal{L}^2(\Omega, \mathcal{E}, \mu)$ of all functions $\Psi : \omega \in \Omega \to \Psi(\omega) \in \mathbb{C}$ that are square-integrable with respect to μ . Every element $A \in \mathcal{A} = \mathcal{L}^\infty(\Omega, \mathcal{E}, \mu)$ is then viewed as an element of $\mathcal{B} = \mathcal{B}(\mathcal{H})$, namely under the identification of the function $A : \omega \in \Omega \mapsto A(\omega) \in \mathbb{C}$ with the multiplication operator $A : \Psi \in \mathcal{H} \mapsto A\Psi \in \mathcal{H}$ where $(A\Psi)(\omega) = A(\omega)\Psi(\omega)$. Under this identification $\mathcal{A}$ becomes a maximal abelian W^*- subalgebra of $\mathcal{B}$; while the center of $\mathcal{B}$, namely $\{C \in \mathcal{B} \,|\, \forall\, B \in \mathcal{B} : [B,C] = 0\}$ is trivial, i.e. consists of the multiples of the identity operator. Note further that every element $B \in \mathcal{B}(\mathcal{H})$ can be viewed as a continuous linear functional on the Banach space $\mathcal{T}(\mathcal{H})$ of all trace-class operators, spanned by the countably additive states; namely $B : T \in \mathcal{T}(\mathcal{H}) \mapsto \mathrm{Tr}\, TB \in \mathbb{C}$; conversely every norm-continuous linear functional on $\mathcal{B}(\mathcal{H})$ obtains in this manner; i.e. $\mathcal{B}(\mathcal{H})$ is the Banach space *dual* of $\mathcal{T}(\mathcal{H})$; equivalently, $\mathcal{T}(\mathcal{H})$ is the *predual* of $\mathcal{B}(\mathcal{H})$. Similarly, the predual of $\mathcal{L}^\infty(\Omega, \mathcal{E}, \mu)$ is the Banach space of $\mathcal{L}^1(\Omega, \mathcal{E}, \mu)$, spanned by the probability distributions which are absolutely continuous with respect to μ.

The interpretation of a quantum state ψ in terms of classical probabilities obtains upon reading (35) separately for each family $\{P_n\}$ of mutually compatible quantum events. The bijective equivalence between the objects described by (35) and (36) is the pragmatic content of Gleason's theorem; cf. e.g. [Emch and Liu, 2002, p. 225]: every quantum state can be uniquely written in the form (36), and every density operator ρ defines through (36) a function ψ satisfying (35), i.e. a quantum state ψ. For the semantic, i.e. the empirical (frequentist vs. subjective) interpretations of states, first in classical probability theories, and then in quantum theories, cf. e.g. [Jaynes, 1967; Emch and Liu, 2002; Emch, 2005]; in particular, see [Uffink, 2006] for the evolution in CSP of the primacy of one over the other of these interpretations of probabilities.

Again in a nutshell, I believe that it serves my purpose well, in most of this essay, to espouse the 'subjective' rather than their 'frequentist' interpretation, namely *to view the state of a physical system* — be it classical or quantum, macroscopic or microscopic — *as a faithful summary of the knowledge one has of the process by which this system has been prepared.* In particular, this semantic view of the quantum state shall translate well from the case of systems with *finitely many degrees of freedom* considered in von Neumann's quantum mechanics, to the systems *with infinitely many degrees of freedom* to be considered in QSP; see subsections 3.4 to 6.3. In particular, while von Neumann's beams or 'ensembles', of independent, identically prepared systems — [von Neumann, 1932c, note 156] — are adequate to describe scattering experiments or the atomic spectroscopy of his time, the view of quantum states that I choose to adopt here accomodates better the description of single macroscopic systems — such as a cup of coffee or a measuring apparatus.

3.2 QSP in von Neumann's formalism

The centerpiece of equilibrum QSP in von Neumann's formalism is the following result [von Neumann, 1932c]:

THEOREM 3. *Let $\mathcal{H}$ be a Hilbert space, H be a self-adjoint operator acting in $\mathcal{H}$ and such that for all $\beta > 0$: the partition function $Z := \mathrm{Tr} \exp(-\beta H)$ be finite. And, with $k > 0$ fixed, let for any state ρ on $\mathcal{B}(\mathcal{H})$*

(37) $S[\rho] = -k\,\mathrm{Tr}\rho \log \rho$.

As H *has discrete spectrum and is bounded below, let* ϵ_o *be its smallest eigenvalue; and let* s *denote either the largest eigenvalue of* H *if* H *is bounded above, or* ∞ *if it is not. Then, for any given* $\epsilon_o < E < s$ *, the maximum of* $S[\rho]$ *, subject to the constraint* $\mathrm{Tr}\rho H = E$ *is reached on the state*

(38) $\rho = Z^{-1}\, e^{-\beta H} \quad \text{with} \quad Z = \mathrm{Tr} e^{-\beta H}$

where the value of β *is determined by the value* E *of the constraint.*

The first part of the proof consists in showing that the maximum occurs on the class of states of the form $\rho = \sum_n \lambda_n\, P_n$ where $\sum_n \epsilon_n\, P_n$ is the spectral resolution of H . After this, the result follows from the classical argument using Lagrange multipliers with respect to the collection of variables $\Lambda = \{\lambda_n\} \subset \mathbb{R}^+$, namely from determining the maximum of the function $S[\Lambda] = -k \sum_n \lambda_n \log \lambda_n$ subject to the simultaneous constraints $\sum_n \lambda_n \epsilon_n = E$ and $\sum_n \lambda_n = 1$.

Note that this variational principle could have been rephrased as defining the state ρ in (38) as the state that minimizes — now under the single constraint $\mathrm{Tr}\rho = 1$, i.e. $\sum_n \lambda_n = 1$ — the Helmholtz free-energy defined as $F := E - TS$ with E and S as in the theorem, and $\beta = kT$ where k is known as the Boltzmann constant (see below).

Note also that, in either of these two forms, this variational principle has its root in the classical statistical physics (CSP) of Boltzmann and Gibbs; cf. [Uffink, 2006]. Conceptually, and very much as in CSP, the von Neumann QSP result involves a consensus on two questions. The first question is to justify the interpretation of S as an entropy. There are two ways to do this.

(i) Firstly, as in CSP, one may identify S with the equilibrium entropy of macroscopic thermal physics upon computing S for well-controlled model(s), such as the ideal gas and finding — for in each of the specific cases considered — that the value of S_{max} obtained through the above theorem coincides with the value of the thermodynamical entropy. It is only at that stage that k may be identified with the universal Boltzmann constant $k \simeq 1.3810^{-23}$ J/deg ; note the units, namely [energy]/[temperature], as is proper for the thermal entropy where T is the integrating factor that allows one to pass from the "heating" differential η to the exact differential $dS = \eta/T$. As fine as that may be for equilibrium CSP/QSP, this identification leaves open the interpretation of S as entropy in non-equilibrium situations.

(ii) The second route to an interpretation of S is to show that $I(\rho) = -S(\rho)$ is a measure of the information content of the state ρ , namely to find empirically meaningful conditions that express the intuitive concept of "information content" and to show that — up to a multiplicative constant — there exists exactly one S that satisfies these conditions. The argument offered by Khinchin [1957] for classical probability distributions involves — *inter alia* — the axiom of consistency under refinements. This argument was transposed to the quantum case by Thirring [1983b] to give:

THEOREM 4. *$S[\rho] = -k\,\mathrm{Tr}\rho\log\rho$ is the only functional satisfying:*

1. *$S[\rho]$ is continuous in ρ, in the sense that it is a continuous function of the eigenvalues of ρ.*

2. *For every finite probability distribution $P = \{p_n \,|\, n = 1, 2, \ldots, N\}$ and every finite collection of states $\{\rho_n \,|\, n = 1, 2, \ldots, N\}$ on a finite collection of Hilbert spaces $\{\mathcal{H}_n \,|\, n = 1, 2, \ldots, N\}$, let ρ be the state defined on $\mathcal{H} = \oplus_{n=1}^{N}\mathcal{H}_n$ by $\rho = \oplus_{n=1}^{N} p_n\rho_n$. One has then: $S[\rho] = S[P] + \sum_{n=1}^{N} p_n S[\rho_n]$ where $S[P]$ is the value of the Khinchin functional for the probability distribution P.*

3. $S\left[\begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix}\right] = k\log 2.$

The first of the three conditions of theorem 4 is clear: an arbitrarily small change in the state should result in an arbitrarily small change in the information it conveys. The second condition expresses refinement under a particular class of partitionings; while the third is only a normalization. Just as in CSP, the *quantum information content* $(-S)$, uniquely specified by these conditions, is formally used to define the *quantum entropy* S.

The second question concerning the conceptual relevance of theorem 3 is to justify the very use of a *variational principle;* compare with [Uffink, 1995]. In my view, for both the classical and the quantum cases, this comes most naturally when one opts for the subjective interpretation of states rather than the frequentist interpretation. Indeed, if one wants the state to account for the knowledge one has of the system, it seems consistent to select for ρ the state that assumes no more information than that expressed explicitly by the constraint.

When the operator H in theorem 3 is taken to represent the energy of the system, the state (38) is called — by analogy to the Gibbs canonical equilibrium state of CSP — the quantum canonical equilibrium state for the natural temperature $\beta = 1/kT$. Note in particular that, in the Schrödinger picture, the evolution generated by H, namely:

$$(39)\quad \forall\, t \in \mathbb{R}: \quad \rho(t) = U(t)\rho U(-t) \quad \text{with} \quad U(t) = \exp^{-i\frac{1}{\hbar}Ht}$$

leaves the canonical equilibrium state invariant, as is to be expected when one wishes to identify the energy-operator with the Hamiltonian of the system.

At first sight, the von Neumann formalism affords a good start for the development of a quantum ergodic theory. To keep things as simple as possible, consider the Hilbert space $\mathcal{L} = \{X \in \mathcal{B}(\mathcal{H}) \,|\, \mathrm{Tr}\, X^*X < \infty\}$ equipped with the scalar product $(X, Y) = \mathrm{Tr}\, X^*Y$. This space is known to mathematicians as the space of Hilbert–Schmidt operators acting on $\mathcal{H}$. In particular, every density matrix is an element of $\mathcal{L}$; and thus this space is also known to physicists as the Liouville space of the quantum system described on $\mathcal{H}$. The advantage of restricting attention to this space is that (39) extends to a unitary action on $\mathcal{L}$:

$$(40)\quad V : (t, X) \in \mathbb{R} \times \mathcal{L} \mapsto V(t)[X] = U(t)XU(-t) \in \mathcal{L} \quad .$$

In the same way as the self-adjoint generator H of the continuous unitary group $\{U(t)|t \in \mathbb{R}\}$ is called the Hamiltonian of the quantum system considered, the self-adjoint generator L of the continuous unitary group $\{V(t)|t \in \mathbb{R}\}$ is called the *Liouvillian* of this system. One has then

THEOREM 5. *Let $H \in \mathcal{B}$ have purely discrete spectrum, i.e. H can be written in the form $H = \sum_n \epsilon_n P_n$ where the P_n are mutually orthogonal projectors adding to I. Then the following limit exists*

$$(41)\quad E_{erg}[X] = \lim_{T\to\infty} \frac{1}{T} \int_0^T dt\, V(t)[X] \quad \text{with} \quad X \in \mathcal{L} \quad ;$$

and

$$E_{erg}[X] = \sum_n P_n X P_n \quad \text{and} \quad \forall\, t \in \mathbb{R}\ : V(t)[E_{erg}[X]] = E_{erg}[X].$$

In particular, the ergodic average $E_{erg}[\rho]$ of a density matrix ρ exists, is again a density matrix, and is time-invariant.

It is tempting to try and consider theorem 5 as proper quantum version of the classical ergodic theorems of Birkhoff [1931] or von Neumann [1932a]. Indeed, the *conclusions* of these classical theorems and of theorem 5 are similar when one reads them in terms of (countably additive) 'states' respectively defined as:

- $A \in \mathcal{L}^\infty(\Omega,\mu) \mapsto \int_\Omega d\mu\, f A \in \mathbb{C}$ where $f \in \mathcal{L}^1(\Omega,\mu)$, f positive with f normalized by $\int_\Omega d\mu\, f = 1$ (for the classical case);
- $A \in \mathcal{B}(\mathcal{H}) \mapsto \mathrm{Tr}\, \rho A \in \mathbb{C}$ where ρ is a density matrice, i.e. a positive trace-class operator with ρ normalized by $\mathrm{Tr}\rho = 1$ (for the quantum case);

and similarly for their respective time-averages.

Note that while the classical theorems are usually followed by a corollary involving the (quasi-)ergodic hypothesis and some discussion of the relevance of the results for the foundations of CSP — for a critical presentation see e.g. [Uffink, 2006, section 6.1] — I do not intend to try and follow suit here, in view of theorems 7 and 8 below which, for the purposes of QSP, cast a shadow on the adequacy of the *assumptions* theorem 5 makes on the Hamiltonian H. For a quantum ergodic theorem better adapted to the needs of QSP, see theorem 25 below.

Nevertheless, two related interesting comments may be made about theorem 5.

(i) If, in this theorem, H is non-degenerate, i.e. if $\forall\, n : \mathrm{dim} P_n = 1$, then $E_{erg}[\rho]$ coincides with

$$(42)\quad Q_o[\rho] = \sum_n \mathrm{Tr}(\rho P_n) P_n = \sum_n (\rho \Psi_n, \Psi_n) P_n$$

where $P_n \Psi_n = \Psi_n$ with $(\Psi_n, \Psi_m) = \delta_{mn}$, and where $Q_o[\rho]$ is thus the density matrix resulting from the von Neumann quantum measuring process [von

Neumann, 1932c, p. 351]; see also subsection 6.3 below. In particular, if ρ is a pure state, i.e. is a projector P_Ψ on some vector $\Psi = \sum_n c_n \Psi_n$, then $Q_o[P_\Psi] = \sum_n |c_n|^2 P_n$ has lost all the information encoded in the relative phases of the coefficients c_n.

(ii) In [von Neumann, 1932c, pp. 380 ff] von Neumann shows that the entropy S of a state does not decrease — and in the generic case does increase — as the result of a measurement, whereas it is constant under the unitary evolution (40). He thus sees in

$$(43)\quad S[Q_o[\rho]] \geq S[\rho]$$

a confirmation that quantum measurements are generically *irreversible* processes. Similarly then, the information encoded in a (non-degenerate) density matrix ρ may only decrease as a result of taking its time-average, a reasonable feature indeed.

Yet, while theorem 5 could have been regarded as the germ of a quantum ergodic theory, the occurrence of monotonic irreversibility in QSP is significantly more elusive, as the next subsection demonstrates.

3.3 Some reasons to go beyond von Neumann's formalism

Some of the problems non-equilibrium QSP has to face are illustrated in a simple spin-lattice model that was originally suggested to me by an actual experiment, the so-called nuclear free-induction relaxation; cf. [Emch and Liu, 2002, section 15.3].

The system consists of a linear chain of N interacting spins $\{\sigma_k = (\sigma_k^x, \sigma_k^y, \sigma_k^n) \,|\, k = 1, \dots, N\}$ with N even (and large, in a sense to be specified later on), and let

$$(44)\quad \sigma_k^x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad , \quad \sigma_k^y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \quad , \quad \sigma_k^z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

be the Pauli matrices acting on $\mathcal{H}_k \simeq C^2$. The Hilbert space of the system is then $\mathcal{H} = \otimes_k \mathcal{H}_k \simeq \mathbb{C}^{2^N}$. In this chain, two spins on sites k and $k+n$ interact with an energy $-J_n \sigma_k^z \sigma_{k+n}^z$, with $J_n > 0$ so that a lower energy is ascribed to configurations in which the $z-$components of these spins are parallel rather than anti-parallel. The whole system is plunged in a homogeneous magnetic field B in the direction z. The total Hamiltonian is chosen to be

$$(45)\quad H_N = -B \sum_{k=1}^{N} \sigma_k^z - \sum_{k=1}^{N} \sum_{n=1}^{N/2} J_n \sigma_k^z \sigma_{k+n}^z \quad \text{with} \quad J_n = 2^{-n} J_o > 0 \ .$$

The system is initially prepared in the state

$$(46)\quad \rho_N = Z_N^{-1} \exp^{-\beta B \sum_{k=1}^{N} \sigma_k^x} \quad \text{with} \quad Z_N = \text{Tr} \exp^{-\beta B \sum_{k=1}^{N} \sigma_k^x} \ .$$

For the three "macroscopic" observables

$$(47)\quad S_N^\alpha = \frac{1}{N}\sum_{k=1}^{N}\sigma_k^\alpha \quad \text{with} \quad \alpha \ \text{standing for} \ x, y, z$$

one computes easily from (39–40) with $H = H_N$ given by (45):

$$(48)\quad \left.\begin{array}{ll} \mathrm{Tr}\,(V_N(t)[\rho_N]\,S_N^x) & = \mathrm{Tr}\,(\rho_N\,S_N^x)\cos(2Bt) f_N(t) \\ \mathrm{Tr}\,(V_N(t)[\rho_N]\,S_N^y) & = \mathrm{Tr}\,(\rho_N\,S_N^y)\sin(2Bt) f_N(t) \\ \mathrm{Tr}\,(V_N(t)[\rho_N]\,S_N^z) & = \mathrm{Tr}\,(\rho_N\,S_N^z) \end{array}\right\}$$

where

$$(49)\quad \left.\begin{array}{c} f_N(t) = f(t)/W_N(t) \quad \text{with} \\ f(t) = \left[\frac{\sin(J_o t)}{J_o\,t}\right]^2 \quad \text{and} \quad W_N(t) = \left[\frac{\sin(2^{-N/2}J_o t)}{2^{-N/2}J_o\,t}\right]^2 \end{array}\right\} .$$

REMARKS 6.

1. For the purpose of discussing the putative irreversibility of the model, the (conservative) Larmor precession $\{\cos(Bt), \sin(Bt)\}$ of the magnetization around the direction z of the magnetic field B is of little or no interest.

2. In favour of the "irreversibility" of the model, one first notes that

   $$(50)\quad \forall\, t \text{ with } |t| << T_N = 2^{N/2}\pi J_o^{-1} \ : \quad f_N(t) \simeq f(t)$$

   and then the decay of $|\mathrm{Tr} V_N(t)[\rho]\,S_N^\alpha|$ is governed by t^{-2}. Therefore, *in this time frame,* the magnetization (48) exhibits an apparent approach to equilibrium.

3. However, against the statement that the model would show an irreversible approach to equilibrium, one observes that

   $$(51)\quad \lim_{t\to T_N} f_N(t) = 1 = f_N(0)$$

   and thus, over the long run the system is periodic in time. This quantum model therefore would seem to confirm the classical Zermelo recurrence objection, or *Wiederkehreinwand*; for the latter, see [Uffink, 2006, section 4.5].

4. The saving grace, nevertheless, is that the period T_N increases exponentially with the size N of the system; see (50). This exponential behaviour is already encountered is CSP, as demonstrated by the Ehrenfest dog-flea model briefly mentioned in subsection 6.1 below. Thus, a modern Galileo would have his *Simplicio* argue that for macroscopically large systems, unaccountable perturbations would set in before T_N is approached, thus irremediably masking this periodicity; compare this to Boltzmann's responses to the Zermelo objection; see again [Uffink, 2006, section 4.5].

5. Upon taking stock of this objection *Salviati* would invoke some modern version of the apocryphal commandment to the effect that "*Thou shalt not interchange limits*" since:

$$(52)\quad \lim_{N\to\infty}\lim_{t\to\infty} f_N(t) \text{ does not exist} \quad \text{but} \quad \lim_{t\to\infty}\lim_{N\to\infty} f_N(t) \text{ exists and is } 0\ .$$

6. The present model presents also a quantum manifestation of the classical Loschmidt reversibility objection, or *Umkehreinwand* — cf. [Uffink, 2006, section 4.3] — as one has, here also:

$$(53)\quad f_N(-t) = f_N(t) \quad \text{and even} \quad f(-t) = f(t)$$

   confirming the classical *Janus* dictum according to which the security of a *postdiction* is the same as that of a *prediction*. Hence, this model indicates that, if the Umkehreinwand were indeed a genuine objection to QSP — which I do not believe it is — the thermodynamical limit would not avoid it, whereas remarks (4) and (5) above show how it may respond to the Wiederkehreinwand.

7. Finally, one serious shortcoming of the present model — not as a model of the particular experiment considered above, but as a model for the approach to equilibrium in a QSP accounting for transport coefficients — is that, even in the limit of $N \to \infty$, the evolution is monitored by an inverse power law in time, rather than an exponential law, as would be required for the type of behaviour encountered in such macroscopic situations as described by Newton's cooling law, Fourier's heat equation, or more generally any macroscopic differential transport equation with linear coefficients.

The model illustrates explicitly some of the essential limitations of the von Neumann formalism for QSP, as manifested in the following two general results. The main assumption of theorems 3 and 5, namely that the Hamiltonian operator H has discrete spectrum, though seemingly innocuous — and actually necessary — when dealing with equilibrium QSP, has one potentially disastrous consequence when one attempts to extend the formalism to non-equilibrium situations: metastases of the classical objections spread into the quantum realm, as we shall now see.

The first result is a quantum version of the classical recurrence theorem of Zermelo. To be mathematically precise recall, in the words of Besicovitch's standard text [Besicovitch, 1954], that a function $f : t \in \mathbb{R} \mapsto f(t) \in \mathbb{C}$ is said to be *almost periodic* in the sense of Harald Bohr, if $f(t+T)$ is approximately equal to $f(t)$ — with an arbitrary degree of accuracy — for infinitely many values of T, these values being spread over the whole real line, in such a way as not to leave empty intervals of arbitrarily great length.

THEOREM 7. *If the Hamiltonian* $H = H^* \in \mathcal{B}(\mathcal{H})$ *has purely discrete spectrum, i.e. if* $H = \sum_n \epsilon_n P_n$ *; and if* $\{V(t) \,|\, t \in \mathbb{R}\}$ *is the unitary action in the Liouville space* $\mathcal{L}$ *defined in (40), then*

(54) $\forall\, X, Y \in \mathcal{L} : f_{X,Y}(t) = \mathrm{Tr}\,(V(t)[X]\,Y)$

is an almost periodic function in t in the sense of H. Bohr.

Proof. $f(t)$ is a Fourier series $\sum_{n,m} a_{n,m} \exp^{-i\frac{1}{\hbar}(\epsilon_n-\epsilon_m)t}$ with $a_{n,m} = \mathrm{Tr}(\,P_n X\, P_m\, y)$; by the Schwartz inequality in $\mathcal{L} : \sum_{n,m} |a_{n,m}|^2$ converges and thus — cf. [Besicovitch, 1954] — f is an almost periodic function of t in the sense of H. Bohr. ■

One might then attempt to get rid of recurrences by assuming — as is certainly allowed in the von Neumann formalism of quantum mechanics, provided $\dim\mathcal{H} = \infty$ — that the spectrum of the Hamiltonian is purely continuous. From the point of view of QSP, however, this cure would raise the following new difficulty, namely that ergodic states may not be countably additive, i.e. may not be representable by density matrices.

To describe this phenomenon, consider the Banach space $\mathcal{B} = \mathcal{B}(\mathcal{H})$ equipped with its usual operator norm; and denote by $\mathcal{B}^*$ its dual, i.e. the Banach space of all continuous, linear functionals on $\mathcal{B}$. Then

(55) $\mathcal{B}^* = \mathcal{A}^* \oplus \mathcal{A}^\perp$

where

i. $\mathcal{A}$ is the space of compact operators on $\mathcal{H}$, i.e. $\mathcal{A} = \{A \in \mathcal{B}(\mathcal{H}) \,|\, \Psi_n \rightharpoonup \Psi \;\Rightarrow\; A\Psi_n \to A\Psi\}$; here, $\rightharpoonup$ and $\to$ respectively denote weak- and strong-convergences in $\mathcal{H}$. When the $^*-$algebra $\mathcal{A}$ is equipped with the operator norm it inherits from $\mathcal{B}(\mathcal{H})$, $\mathcal{A}$ is closed in $\mathcal{B}(\mathcal{H})$ and thus is a Banach space on its own; in fact $\mathcal{A}$ is the only non-trivial closed two-sided $^*-$ideal of $\mathcal{B}$.

ii. For every $\varphi \in \mathcal{A}^*$, the dual of $\mathcal{A}$, there exists a unique trace-class operator $R \in \mathcal{T} = \{B \in \mathcal{B} \,|\, \mathrm{Tr}\,(B^*B)^{\frac{1}{2}} < \infty\}$ such that $\forall\, A \in \mathcal{A} :\; \varphi(A) = \mathrm{Tr}(R\,A)$. In particular, to every positive, continuous linear functional ψ on $\mathcal{A}$ such that $\sup_{A\in\mathcal{A},\|A\|\leq 1} \|\psi(A)\| = 1$ there corresponds a unique density matrix, and conversely.

iii. $\mathcal{A}^\perp = \{\varphi \in \mathcal{B} \,|\, A \in \mathcal{A} \Rightarrow \varphi(A) = 0\}$.

Note that each of the inclusions $\mathcal{T} \subseteq \mathcal{L} \subseteq \mathcal{A} \subseteq \mathcal{B}$ is strict iff $\mathcal{H}$ is infinite-dimensional, a condition that is required whenever one wants to avoid recurrences, since $\dim\mathcal{H} < \infty$ obviously entails that the spectrum of H is purely discrete, and then theorem 7 applies.

We can now make precise the above mentioned difficulty concerning the description of ergodic states within the context of countably additive states:

THEOREM 8. *Let* $H \in \mathcal{B}(\mathcal{H})$ *be the self adjoint generator of any strongly continuous unitary group* $\{U(t) \mid t \in R\}$ *acting on* $\mathcal{H}$; *and, with* t *running over* $\mathbb{R}$, *let* $\rho \in \mathcal{T} \mapsto \rho(t) = U(t)\rho U(-t) \in \mathcal{T}$ *describe the evolution of any density matrix* ρ; *further, let* ψ_t *denote the corresponding (countably additive!) state* $\psi_t : B \in \mathcal{B}(\mathcal{H}) \mapsto \psi_t(B) = \mathrm{Tr}(\rho(t)B) \in \mathbb{C}$. *Then, it follows that:*

a. For every compact observable $A \in \mathcal{A}$ the ergodic limit

$$(56)\quad \lim_{T\to\infty} \frac{1}{T}\int_o^T dt\ \psi_t(A)$$

exists and defines a positive linear functional $E_\infty[\psi]$ on $\mathcal{A}$.

b. If, moreover, the spectrum of H is purely continous, then $E_\infty[\psi]$ cannot be extended to a countably additive state on $\mathcal{B}(\mathcal{H})$.

Proof. For the part (a), the economical strategy is to take advantage of two density theorems, namely: (i) when $\mathcal{L}$ is equipped with its Hilbert-Schmidt norm, it contains $\mathcal{T}$ as a dense subspace; and (ii) when $\mathcal{A}$ is equipped with the operator norm it inherits from $\mathcal{B}(\mathcal{H})$, it contains $\mathcal{L}$ as a dense subspace. Hence, one can uniquely lift the evolution from $\mathcal{T}(\mathcal{H})$ to a unique unitary evolution on the Hilbert space $\mathcal{L}$ where one can use the classical ergodic theorem — [von Neumann, 1932a], or [Emch and Liu, 2002; Uffink, 2006] — to assert the existence of the ergodic limit for any pair $(X, Y) \in \mathcal{L}\times\mathcal{L}$, and in particular for any pair $(\rho, A) \in \mathcal{T}\times\mathcal{L}$. Upon recalling the duality $\mathcal{A} = \mathcal{T}^*$, the ergodic result is then extended by continuity from $\mathcal{T}\times\mathcal{L}$ to $(\rho, A) \in \mathcal{T}\times\mathcal{A}$.

To prove part (b), one notices that, on the one hand, this limit is given, for every $A \in \mathcal{A}$ by $E_o[\psi](A) = \mathrm{Tr}(\sum_n (P_n\, \rho\, P_n\, A)$ where $\{P_n\}$ is the set of all the projectors corresponding to the discontinuous jumps in the spectral family of H. Hence, when H has continuous spectrum, this set is empty and thus $\forall\, A \in \mathcal{A} :\ E_o[\psi](A) = 0$. On the other hand, $E_o[\psi]$ certainly extends further than $\mathcal{A}$; for instance the ergodic limit obviously exists for any $X \in \{H\}'$, i.e. for all bounded observables that are constants of the motion; in particular $E[\psi_o](I) = 1$. Hence, even if $E[\psi_o]$ could be extended to a state on $\mathcal{B}$, this state would belong to $\mathcal{A}^\perp$ and thus would have no countably additive component in the direct sum decomposition (55). ■

The proof of the theorem shows that the same objection can be raised with any Hamiltonian the spectrum of which contains even only one interval of continuity. Taking Hamiltonians that are still self-adjoint, but not bounded above would only raise more technical problems without providing a solution to the basic limitation exposed in theorem 8.

Hence, von Neumann's formalism for QSP leads non-equilibrium QSP into the horns of a dilemma: either the evolution is almost periodic or the ergodic states are not countably additive. In particular, non-equilibrium states cannot approach asymptotic states that can be described by density matrices.

To make a bad situation even worse, Zeh discovered — admittedly, long after von Neumann's treatise had appeared and yet relevant to the thrust of this section — that there are serious empirical difficulties with the concept of an isolated quantum system [Zeh, 1970; Wigner, 1984]. Could certainty be fading out? [Prigogine, 1997]. Zeh's original observation has led to the development of the concept of *decoherence*; cf. Landsman [Landsman, 2006, section 7.1]. I very briefly address this and some related issues in section 6 below.

Even in equilibrium QSP, the anchor provided by von Neumann was slipping: the formalism cannot account for the coexistence of thermodynamical phases; for a response to this objection, see subsection 5.7 below.

In counterpoint to these questions, one fundamental problem needs to be addressed: namely that the von Neumann formalism is not adequate to describe typical many-body systems where an infinite number of degrees of freedom are brought into the picture. The prescribed remedy is discussed in the next subsection.

3.4 Haag–Kastler's axioms and Takeda's inductive limits

This subsection outlines a formalism proposed to deal effectively with the non-relativistic many-body problems in QSP. This formalism was born out of the axiomatic responses prompted by the diagnosis of a mid-life crisis in relativistic Quantum Field Theory [QFT] [van Hove, 1952; Friedrichs, 1953; Wightman and Schweber, 1955]; the nail in the coffin was driven by Haag [1955]; cf. e.g. the famous Haag theorem and its embalmings in [Barton, 1963, section 14], [Streater and Wightman, 1964, section IV.5], and/or [Emch, 1972a, section 3.d]. The algebraic axiomatization is presented here with sufficiently elementary details, yet with enough restraint to eschew the "imperialistic" label sometimes attached to it.

The main idea is to account for the local structure of infinitely extended systems. In their original proposal, Haag and Kastler [1964] mention several precedents in axiomatic QFT; among these [Haag, 1959a; Haag and Schroer, 1962]; see also [Haag, 1959b]. (I first heard of the algebraic approach in seminars in Geneva, where Araki presented some aspects of his Zurich lectures [Araki, 1961/2].) Segal's early advocacy of an algebraic approach [Segal, 1947] ought also to be mentioned.

This subsection is divided in two complementary parts: the first part presents a description of the general structure; the second illustrates this structure with an example, the 1-dimensional quantum spin-lattice.

Part I. The general structure.

One begins by selecting an absorbing directed net $\mathcal{F}$ of regions Λ of finite extension in space; usually, the space is the Minkowski space $\mathbb{M}^{n+1}$ for relativistic QFT, the Euclidean space $\mathbb{R}^n$ or a lattice $\mathbb{Z}^n$ for non-relativistic QSP. The case of immediate physical interest is $n = 3$, but exploratory models are often constructed with $n = 1, 2$. Recall that a directed net is a partially ordered set — here the order relation is the usual set-theoretical inclusion — such that for every pair of elements Λ_1, Λ_2 in $\mathcal{F}$ there is at least one element $\Lambda \in \mathcal{F}$ such that $\Lambda_1 \subseteq \Lambda$ and $\Lambda_2 \subseteq \Lambda$. To say that this net is absorbing is to say that for every point x in space there exists at least one element $\Lambda \in \mathcal{F}$ such that $x \in \Lambda$. The symbol $\Lambda_1 \bowtie \Lambda_2$ will be used to signify that two regions Λ_1 and Λ_2 are causally disjoint, i.e. in QFT, these regions are spacelike to one another; and in non-relativistic QSP, they are disjoint in the set-theoretical sense, i.e. $\Lambda_1 \cap \Lambda_2 = \emptyset$. G denotes a group of rigid motions in the space, namely the inhomogeneous Lorentz group for $\mathbb{M}^{n+1}$; the Euclidean group for $\mathbb{R}^n$; or the group of lattice translations for $\mathbb{Z}^n$.

Secondly, to every $\Lambda \in \mathcal{F}$ one assigns a C^*−algebra $\mathcal{A}_\Lambda$; without loss of generality, one may assume that $\mathcal{A}_\Lambda$ has an identity I_Λ. This assignment is subject to the following three postulates.

POSTULATE 9 (Isotony). Whenever $\Lambda_1 \in \mathcal{F}$ and $\Lambda_2 \in \mathcal{F}$ satisfy $\Lambda_1 \subseteq \Lambda_2$, one is given an injective *−homomorphism $i_{21} : \mathcal{A}_{\Lambda_1} \to \mathcal{A}_{\Lambda_2}$ such that

1. $i_{21}(I_{\Lambda_1}) = I_{\Lambda_2}$
2. $\Lambda_1 \subseteq \Lambda_2 \subseteq \Lambda_3 \;\Rightarrow\; i_{32} \circ i_{21} = i_{31}$.

The following result was proven by Takeda [1955].

THEOREM 10. *Let $\mathcal{F}$ be a directed net, and $\{\mathcal{A}_\Lambda \mid \Lambda \in \mathcal{F}\}$ satisfy the isotony postulate. Then there exist: a C^*−algebra $\mathcal{A}$ with unit I; and a family of injective $*$−homomorphisms $\{i_\Lambda : \mathcal{A}_\Lambda \to \mathcal{A} \mid \Lambda \in \mathcal{F}\}$ such that*

1. $\forall\, \Lambda \in \mathcal{F} \;:\; i_\Lambda(I_\Lambda) = I$;
2. $\Lambda_1 \subseteq \Lambda_2 \;\Rightarrow\; i_{\Lambda_1}(\mathcal{A}_{\Lambda_1}) \subseteq i_{\Lambda_2}(\mathcal{A}_{\Lambda_2})$;
3. $\bigcup_{\Lambda\in\mathcal{F}} i_\Lambda(\mathcal{A}_\Lambda)$ *is a norm-dense sub−*algebra of* $\mathcal{A}$.

The C^*−algebra $\mathcal{A}$ is called the C^**−inductive limit* of the net $\{\mathcal{A}_\Lambda \mid \Lambda \in \mathcal{F}\}$. We will use hereafter the notations

$$\mathcal{A}_o := \bigcup_{\Lambda\in\mathcal{F}} i_\Lambda(\mathcal{A}_\Lambda) \quad \text{and} \quad \mathcal{A} = {}^n\overline{\mathcal{A}_o}.$$

POSTULATE 11 (Local commutativity). Whenever $\Lambda_1, \Lambda_2 \in \mathcal{F}$ satisfy $\Lambda_1 \bowtie \Lambda_2$, and $\Lambda_3 \in \mathcal{F}$ is such that both $\Lambda_1 \subseteq \Lambda_3$ and $\Lambda_2 \subseteq \Lambda_3$:

$$A_1 \in \mathcal{A}_{\Lambda_1} \text{ and } A_2 \in \mathcal{A}_{\Lambda_2} \;\Rightarrow\; i_{31}(A_1)\, i_{32}(A_2) = i_{32}(A_2)\, i_{31}(A_1) \quad .$$

The following result is then immediate.

COROLLARY 12. *If* $\Lambda_1 \bowtie \Lambda_2$, *then*

$$A_1 \in \mathcal{A}_{\Lambda_1} \text{ and } A_2 \in \mathcal{A}_{\Lambda_2} \;\Rightarrow\; i_{\Lambda_1}(A_1)\, i_{\Lambda_2}(A_2) = i_{\Lambda_2}(A_2)\, i_{\Lambda_1}(A_1) \quad .$$

For the aspects of QSP considered here, it will be an innocent abuse of language to refer to the above postulate as simply the *postulate of locality.*

POSTULATE 13 (Covariance). An action $\nu : (g, A) \in G \times \mathcal{A}_o \to \nu_g[A] \in \mathcal{A}_o$ is given so that for every region $\Lambda \in \mathcal{F}$, ν_g induces a $^*-$isomorphism between $\mathcal{A}_\Lambda$ and $\mathcal{A}_{g[\Lambda]}$, where $g[\Lambda]$ denotes the image of the region Λ under the point transformation g.

Upon using theorem 10, this can be lifted to $\mathcal{A}$, namely $\nu_g[i_\Lambda(\mathcal{A}_\Lambda)] = i_{g[\Lambda]}(\mathcal{A}_{g[\Lambda]})$:

COROLLARY 14. *The action of G extends by continuity to a norm-continuous group representation $\nu : g \in G \to \mathrm{Aut}(\mathcal{A})$.*

DEFINITION 15. With the above notations, $\mathcal{A}_o$ is called the algebra of local observables; and $\mathcal{A}$ is called the algebra of quasi-local observables. Moreover, if φ is a state on $\mathcal{A}$ such that $\forall\, g \in G : \varphi \circ \nu_g = \varphi$, let π_φ be the corresponding GNS representation. The von Neumann algebra $\mathcal{N}_\varphi = \pi_\varphi(\mathcal{A})''$ is called the algebra of global observables relative to the state φ.

Note that quasi-local observables involve norm limits; they are therefore general, algebraic objects that can be defined abstractly, i.e. without reference to any particular Hilbert space representation. In contrast, global observables that are not quasi-local involve weak-operator limits, and thus depend on the Hilbert space representation in which these limits are taken; for the purposes of QSP these observables depend, via the GNS construction, on the physical situation for which they are defined, i.e. on the state with respect to which they are considered. This aspect of the theory will be discussed in details in subsection 3.5 — see in particular the preliminaries to scholium 23, and remark 26(1) — and it will be essential for the treatment of phase transitions, inasmuch as these involve averages of observables, for instance the spontaneous magnetization in ferromagnets; see subsection 5.7 below.

Part II. A concrete example of a net of observable-algebras.
This example exhibits the construction of the algebra of observables for an infinite quantum spin-lattice system that obtains in the thermodynamical limit of finite systems such as the one treated in subsection 3.3 above. Consider indeed an infinite 1-dimensional lattice $\mathbb{Z}$ with a quantum $\frac{1}{2}-$spin sitting at each node (or "site"); hence a copy $\mathcal{A}_k$ of the C^*-algebra $\mathcal{M}(2,\mathbb{C})$ of 2×2 matrices with complex entries is associated to each site $k \in Z$; i.e. $\mathcal{A}_k$ is generated by the Pauli matrices (44), i.e. by the three observables corresponding to the three components of a $\frac{1}{2}-$spin sitting at site k.

Let now $\mathcal{F}$ be the net of all finite subsets $\Lambda \subset Z$. To each of these Λ is then associated the "local" C^*-algebra $\mathcal{A}_\Lambda = \otimes_{k\in\Lambda}\mathcal{A}_k$ which is thus a copy of $\mathcal{M}(2^{|\Lambda|},\mathbb{C})$, where $|\Lambda|$ denotes the number of sites in Λ.

Let now Λ_1 and Λ_2 be two finite regions, with $\Lambda_1 \subseteq \Lambda_2$. An injective $^*-$homomorphism of $\mathcal{A}_{\Lambda_1}$ into $\mathcal{A}_{\Lambda_2}$ obtains by linearity from its restriction to monomials, namely

$$i_{21}(A_1 \otimes A_2 \otimes \ldots \otimes A_{|\Lambda_1|}) \;=\; B_1 \otimes B_2 \otimes \ldots \otimes B_{|\Lambda_2|}$$

with

$$\forall\, k \in \Lambda_2 \;:\; B_k = \begin{cases} A_k & \text{if} \quad k \in \Lambda_1 \\ I_k & \text{if} \quad k \notin \Lambda_1 \end{cases} \quad .$$

These inclusions satisfy postulate 9.

Here, two finite regions are in the relation $\Lambda_1 \bowtie \Lambda_2$ whenever $\Lambda_1 \cap \Lambda_2 = \emptyset$. Since the commutators of observables attached to individual sites vanish whenever the two sites are different, any two observables attached to disjoint regions do commute. Formally, this is to say that postulate 11 is satisfied.

Finally, let $G := \mathbb{Z}$ denote the additive group of translations of the lattice. To define the action of G on the algebra of local observables it is sufficient to notice that for all $g \in G$ and all $\Lambda \in \mathcal{F} \,:\, |g[\Lambda]| = |\Lambda|$, so that both $\mathcal{A}_{g[\Lambda]}$ and $\mathcal{A}_\Lambda$ are copies of the same matrix algebra, namely $\mathcal{M}(2^{|\Lambda|}, \mathbb{C})$: the images of a local observable and its translate are simply different copies of the same matrix; this indeed defines ν_g in such a way that postulate 13 is satisfied.

3.5 Quantum ergodic theory and macroscopic observables

While classical ergodic theory concerns itself with *measures* μ that are invariant under a group G and their *mixing* properties, quantum ergodic theory discusses the properties of $G-$invariant *states* and their *clustering* properties. Accordingly, in this subsection I will discuss also the roles of *space and/or time averages* in explaining at least part of the success of QSP; compare with [Uffink, 2006] and, in particular, with [Earman and Rédei, 1996].

Therefore, one question to be addressed in this subsection must be whether and how ergodic theory may serve as a cornerstone to build up statistical mechanics. Traditionally, under the impetus of the emphasis the Ehrenfests' placed on Boltzmann's ergodic hypothesis (or rather its measure-theoretical version, the quasi-ergodic hypothesis), the group G is taken to be the group $\mathbb{R}$ governing the *time evolution.* Nevertheless, partly in response to some swaying to and fro in Boltzmann's own writings, the jury is still out on the fundamental issue of this hypothesis' relevance for the foundations of CSP; see in particular [Uffink, 2006, section I.3, and subsections I.4.3 and I.6.1]. It is also remarkable that already Gibbs [Gibbs, 1902] chose to emphasize the role of mixing properties, i.e. properties which are stronger than metric transitivity and make more precise the presumption that the dynamics is 'erratic'; cf. e.g. [Uffink, 2006, sections I.4.1 and I.5]; other issues are touched upon in [Emch and Liu, 2002, pp. 317–330]; and, for a pertinent account that takes stock of the work done in the second half of the twentieth century, by the Lebowitz and Sinai schools, see [Szasz, 1996].

Consequently, I will concentrate here on two sub-questions: (i) the extent to which the mathematics of classical ergodic theory may be generalized to the formalism of quantum theory; and (ii) the extent to which such generalizations may help formulate better certain aspects of the foundations of QSP.

The answer to the first of these sub-questions is that much of the mathematics goes through, with some minor adjustments. The answers to the second is more complex. On the one hand, as long as the focus remains on the time evolution the main issues persist, among which is the paucity of realistic models. On the other hand, when the group G deals with the geometry of the problem, quantum ergodic theory — especially and the roles of averages, and theorems linking extremal invariance and clustering properties — does help distinguish, within QSP, the quantum aspects of the microscopic description and the classical aspects of the macroscopic world. Thus, I divide the presentation in two parts, according to whether ergodicity is considered with respect to the time evolution or with respect to space symmetries.

A. Ergodicity with respect to time

Some insight is gained from a model first proposed in [Ford *et al.*, 1965], which comes in two versions, classical and quantum. The quantum version has been controlled mathematically in [Davies, 1972]. It is proven there that an infinite 1-dimensional chain of weakly coupled 1-dimensional quantum harmonic oscillators may serve as a thermal reservoir for a single 1-dimensional quantum oscillator in the chain and that a diffusion equation governs the evolution of the latter. This is accomplished by a rigorous treatment of the van Hove weak-coupling/long-time limit about which more will be said in subsection 6.1 below. Anticipating some mathematical definitions to be made precise later on — see paragraph 5.3.C — it is sufficient for the present discussion to register that, in this van Hove limit, a reduced evolution obtains which is a contractive semi-group of completely positive maps $\{\gamma_s \mid s \in \mathbb{R}^+\}$ acting on the von Neumann algebra $\mathcal{N}_o \simeq \mathcal{B}(\mathcal{L}^2(\mathbb{R},\mathbb{C}))$ attached to the site of the single oscillator considered. Moreover this evolution, when observed from any one-dimensional subspace $\{x\mathbf{u} \mid x \in \mathbb{R}\}$ in the 2-dimensional phase space $\{\xi P + \eta Q \mid \zeta = (\xi,\eta) \in \mathbb{R}^2\}$ of the single oscillator, is described by a classical distribution $\mu(x,s)$ that satisfies for all $s \in \mathbb{R}^+$ the *diffusion equation:*

$$(57)\quad \partial_s \mu(x,s) = D\,\left[\,\partial_x^2 + \beta(V'(x)\,\partial_x + V''(x))\,\right]\,\mu(x,s)$$

where $\beta = 1/kT$ is the natural temperature, $V = \frac{1}{2}\omega\, x^2$ is a harmonic potential, while the diffusion constant D and the frequency ω are numbers, the values of which depend only on the direction $\zeta/|\zeta| \in \mathbb{R}^2/S^1$. Note that the corresponding invariant measure is the canonical equilibrium, Gaussian measure $\mu(x) = Z^{-1}\,\exp(-\beta V(x))$ with $Z = \int_R dx \mu(x)$, i.e. $Z^{-1} = \sqrt{2\pi\beta\omega}$.

The point of the model here is that the dissipative system described by the contractive *semi-group* $\{\gamma_s \mid s \in \mathbb{R}^+\}$ governing this Markovian diffusion process admits a canonical dilation to a conservative dynamical system. Indeed, there exists a *group* $\{\alpha_s \mid s \in \mathbb{R}\}$ of automorphisms of the von Neumann algebra $\mathcal{N} = \pi_\varphi(\mathcal{A})''$ describing the full chain of oscillators in the equilibrium state φ corresponding to the temperature β when the interactions are switched off. In conformity with subsection 3.4 the algebra of quasi-local observables $\mathcal{A}$ is here the C^*-algebra $\otimes_{k\in\mathbb{Z}}\mathcal{N}_k$ where the $\mathcal{N}_k$ are copies of $\mathcal{N}_o$. The free equilibrium state has the form $\varphi = \otimes\varphi_k$ where φ_k is the von Neumann canonical equilibrium state for the oscillator at the site k. Let now i be the injection of $\mathcal{N}_o$ into $\mathcal{N}$ and φ_o

denote the restriction of φ to $\mathcal{N}_o$, i.e. $\forall N_o \in \mathcal{N}_o : \varphi_o(N_o) = \varphi(i[N_o])$. Let further $E : \mathcal{N} \to \mathcal{N}_o$ be the canonical conditional expectation with respect to the state φ, satisfying $\varphi_o \circ E = \varphi$. The sense in which $\{\mathcal{N}, \alpha, E\}$ is a dilation of $\{\mathcal{N}_o, \gamma\}$ is that

$$(58) \quad \forall (s, N_o) \in \mathbb{R}^+ \times \mathcal{N}_o : \gamma_s[N_o] = E \circ \alpha_s \circ i[N_o] \quad .$$

For details, see [Emch, 1976] where, in particular, this result was noted to be very reminiscent of the classical flow of Brownian motion constructed by Hida [Hida, 1970] who also proved that this flow is a classical Kolmogorov flow, in the sense of the following definition.

DEFINITION 16. A classical dynamical system $\{\Omega, \mathcal{E}, \mu, \alpha^*\}$ consisting of a probability space $\{\Omega, \mathcal{E}\}$, a probability measure μ, and a group $\{\alpha^*{}_t \mid t \in \mathbb{R}\}$ of automorphisms of $\{\Omega, \mathcal{E}\}$ such that $\forall t \in \mathbb{R} : \mu \circ \alpha^*{}_t = \mu$, is said to be a classical Kolmogorov flow whenever there exists a $\sigma-$subring $\mathcal{A} \subset \mathcal{E}$ such that, with the notation $\mathcal{A}_t = \alpha^*{}_t[\mathcal{A}]$:

(1) $\forall t > 0 : \mathcal{A} \subset \mathcal{A}_t$; (2) $\bigvee_{t \in \mathbb{R}} \mathcal{A}_t = \mathcal{E}$; and (3) $\bigwedge_{t \in \mathbb{R}} \mathcal{A}_t = \{\emptyset, \Omega\}$.

Kolmogorov flows are characterized among classical dynamical systems by their having strictly positive dynamical entropy; thus they sit pretty high in the classical ergodic hierarchy, above the Lebesgue spectrum condition, and thus above the weaker conditions of mixing and ergodicity; for didactic accounts, cf. e.g. [Arnold and Avez, 1968; Cornfeld *et al.*, 1982].

The conservative quantum dynamical system described above as the canonical dilation of a contractive semigroup, does satisfy a quantum generalization of definition 16, namely:

DEFINITION 17. A quantum dynamical system $\{\mathcal{N}, \varphi, \alpha\}$ consisting of a von Neumann algebra $\mathcal{N}$, a faithful normal state φ on $\mathcal{N}$, and a group $\alpha = \{\alpha_t \mid t \in \mathbb{R}\}$ of automorphisms of $\mathcal{N}$, with $\forall t \in \mathbb{R} : \varphi \circ \alpha_t = \varphi$, is said to be a generalized Kolmogorov flow whenever there exists a von Neumann subalgebra $\mathcal{A} \subset \mathcal{N}$ such that, with the notation $\mathcal{A}_t = \alpha_t[\mathcal{A}]$:

(1) $\forall t > 0 : \mathcal{A} \subset \mathcal{A}_t$; (2) $\bigvee_{t \in \mathbb{R}} \mathcal{A}_t = \mathcal{N}$; (3) $\bigwedge_{t \in \mathbb{R}} \mathcal{A}_t = \mathbb{C}I$; and

$$(4) \ \forall t \in \mathbb{R} : \tau_t[\mathcal{A}] = \mathcal{A},$$

where $\{\tau_t \mid t \in \mathbb{R}\}$ is the modular group canonically associated to φ.

REMARKS 18.

1. The $\bigvee$ in condition (2) involves a weak-operator closure, namely (2) means that $\mathcal{N}$ is the smallest von Neumann algebra that contains all the $\mathcal{A}_t$; the $\bigwedge$ in condition (3) is simply the usual intersection; thus (3) signifies that no operator belongs to all $\mathcal{A}_t$ unless it is a multiple of the identity.

2. The modular group τ will be introduced in section 4; let it suffice to say here that, if we were dealing with a finite system, τ would be the group of automorphisms of $\mathcal{N}$ associated to the Hamiltonian corresponding to von Neumann's canonical equilibrium density matrix.

3. Definition 17 encompasses definition 16 when $\mathcal{N}$ is taken to be the abelian von Neumann algebra $\mathcal{L}^\infty(\Omega, \mathcal{E})$ acting on the Hilbert space $\mathcal{H} = \mathcal{L}^2(\Omega, \mathcal{E}, \mu)$; in this case $\forall\, t \in R\, :\; \tau_t = id$, and condition (4) is then trivially satisfied.

4. In the general case, condition (4) is necessary to ensure the existence of a conditional expectation $E : \mathcal{N} \to \mathcal{A}$.

5. Except for the positivity of the dynamical entropy — which depends on a consensus that is still pending about a physically meaningful definition of quantum dynamical entropy; see nevertheless [Narnhofer and Thirring, 1994b; Tuyls, 1998] and references therein — all the ergodic properties of classical Kolmogorov systems carry over straightforwardly from the classical to quantum realm [Emch, 1976]. In the model described above these properties are exhibited in the quantum triple $\{\mathcal{N}, \varphi, \alpha\}$.

6. Definition 17 was first proposed in [Emch, 1976]. Generalizations of this definition, involving the passage from W$^*-$ to C$^*-$algebras, were then explored in [Narnhofer and Thirring, 1989].

7. The material of the present remark may be found in [Arnold and Avez, 1968] and is inserted here only as a preparation for the next remark. In classical ergodic theory the next rung up the ergodic ladder, just above Kolmogorov flows, is occupied by Anosov flows. These flows formalize an observation made in 1898 by Hadamard, namely that the geodesics on manifolds of negative curvature exhibit exponential sensitivity to initial conditions, in contrast with the usual linear sensitivity characteristic of free flows on flat manifolds. If the manifold is furthermore compact, one may intuitively expect that Hadamard's observation entails some kind of mixing behaviour. This is indeed the case: the first ever Hamiltonian flow shown to be ergodic — the geodesic flow on a compact surface of constant negative curvature — is already an Anosov flow. These flows exhibit exponentially contracting and expanding directions transversal to the direction of the flow, thus prefiguring a microscopic explanation for the empirically observed Lyapunov coefficients. The discrete-time archetype is the Arnold CAT map operating of the torus $T^2 := \mathbb{R}^2/\mathbb{Z}^2$. One ought to note that up to Kolmogorov flows, classical ergodic theory may be viewed as a chapter in probability theory; Anosov flows, in addition, involve an essential appeal to differential geometry, as was recognized only in the second half of the twentieth century through the work of the Russian school.

8. In order to explore possible quantum extensions of the concept of Anosov flow, a quantum analog of the latter has been devised by the present author in collaboration with Narnhofer, Sewell and Thirring [Emch *et al.*, 1994a]; for an antecedent, see [Benatti *et al.*, 1991a]; for a discussion of dynamical entropy in this context, see [Andries *et al.*, 1995]; for reviews and some general perspectives, see [Narnhofer, 2001; Narnhofer, 2005].

One essential feature of this extension is that now the phase space of this quantum CAT map is the noncommutative torus T_θ^2 , an ubiquitous staple of Connes' noncommutative geometry; cf. e.g. [Connes, 2000, section XIII] or [Garcia–Bondia *et al.*, 2003, chapter 12]; and for the place these tori occupy in the geometric quantization programme, cf. [Emch, 1998b]. As for quantum ergodic theory, it was noted already in [Emch *et al.*, 1994a] that the generators of the expanding and contracting horocycles form a basis in the 2-dimensional distinguished space of derivations that are not approximately inner — i.e. cannot be uniformly approximated by inner derivations [Garcia–Bondia *et al.*, 2003, section 12.3].

The presence of expanding and contracting directions in quantum as well as in classical Anosov flows offers a bridge from classical to quantum chaos. The problem of what is quantum chaos — or what it ought to be — has received attention from different prospectives; cf. e.g. [Gutzwiller, 1990]; for a philosophical perspective, cf. [Belot and Earman, 1997], and for a recent review, cf. in this volume [Landsman, 2006, section 5.6].

The investigations sketched in this remark, with applications to QSP in view, also have a mathematical parallel in QFT, cf. [Borchers, 1999; Wiesbrock, 1997]; see also subsection 5.5 below.

Summary and warning. It seems fair to infer that the *mathematical* generalization of classical ergodic kinematics to the quantum realm will carry through quite well. Nevertheless, the discussion of the underlying *physical* dynamics, when confronted with Hamiltonian mechanics, does not fare any more smoothly in the quantum case than it does in the classical case. Some of the conceptual problems may already be illustrated with the help of the model discussed at the beginning of this subsection. There, the dissipative dynamical system $\{\mathcal{N}_o, \gamma\}$ may be viewed as the reduced dynamics of two different conservative dynamical systems; both of these act on the same infinite assembly of harmonic oscillators. From the first system the reduced dynamics obtains only through the van Hove limit which compounds the very long-time effects — on a single subsystem — of a very weak coupling with, and within, the bath. But there is nothing in common between the time scale of the dynamics that governs the original conservative system and the time scale pertaining to the other conservative system, viz. the one obtained as the canonical dilation of the dissipative system. So there is little reason to believe that the ergodic behaviour of the latter reflects any global dynamical property of the former.

While this may be blamed on some naive modeling, it nevertheless emphasizes that the time scale of the conservative microscopic description and that of the emerging macroscopic description may differ significantly. In more sophisticated models, this will have to be taken into consideration and the complicated behaviour of the microscopic description may have to be washed away — one way or another — before a clean ergodic behaviour is manifested at the macroscopic level. It appears that van Hove's idea is a reasonable way to do this; see subsection 6.1

below.

Starting with their initial motivation in Boltzmann's works, most presentations of classical ergodic theory focus on the properties of the time-evolution, in particular on the transitivity of measures and the *time*-averages of observables. Its generalization to the quantum realm invites the consideration of other aspects of classical ergodic theory, namely the *space* averages with respect to the actions of *other* groups beside those that govern the evolution. This will be done in the second part of this subsection.

B. Ergodicity with respect to space

As was already recognized by Haag [1959b] for QFT, the "other" group of most immediate relevance to QSP is the group of space translations, introduced as a part of the postulate of covariance in the Haag–Kastler axioms; cf. postulate 13 above. With $n = 1, 2, \ldots$, let $\mathbb{X}^n$ denote either the Euclidean space $\mathbb{R}^n$ or the "cubic" lattice $\mathbb{Z}^n$; and let $|x|$ denote the length of the vector $x \in \mathbb{X}^n$. Henceforth, we concentrate on the abelian group $G \simeq \mathbb{X}^n$ of all translations $x \in \mathbb{X}^n \mapsto x + a \in \mathbb{X}^n$ where $a \in \mathbb{X}^n$. Let further $\{\mathcal{A}_\Lambda \mid \Lambda \in \mathcal{F}\}$ be the corresponding Haag–Kastler net of local algebras, and $\mathcal{A}$ be their C^*- inductive limit, with $\mathcal{A}$ equipped with the group of automorphisms $\{\nu_a \mid a \in X^n\}$ defined as in corollary 14. Let again $\mathcal{A}_o \subset \mathcal{A}$ denote the algebra of local observables. For any fixed pair (Λ_1, Λ_2) of elements in $\mathcal{F}$, there exists $a_{12} \in G$ such that $a[\Lambda_1] \bowtie \Lambda_2$ for all $a \in G$ with $|a| > |a_{12}|$. Consequently, by locality (see postulate 11) whenever $a \in G$ with $|a| > |a_{12}|$, $A_1 \in \mathcal{A}_1$ and $A_2 \in \mathcal{A}_2$, we have $\nu_a[A_1]\, A_2 = A_2\, \nu_a[A_1]$. By continuity, this entails

COROLLARY 19. *For all* $A, B \in \mathcal{A}$: $\lim_{|a|\to\infty} \| \nu_a[A]\, B - B\, \nu_a[A] \| = 0$, *i.e. the group* G *of translations acts on the algebra* $\mathcal{A}$ *of quasi-local observables in a* norm-asymptotic abelian manner.

This property makes no sense in the original von Neumann framework for the quantum mechanics of finite systems. In the generalized Haag–Kastler framework devised for infinite systems, this statement which is straighforwardly correct for space translations is rarely satisfied by the time evolution in realistic models that have been controlled.

This raises three questions: the first is whether this property has useful consequences; the second is whether this property can be weakened without jeopardizing the consequences that may be derived from it; and the third is whether any of the weakened forms of this property may be satisfied by the time evolution. I will argue that the answers to the first two questions are "yes". Specifically, in regard to the first question, see in particular corollary 30 below; and in response to the second, see the forthcoming theorem 25. However, here again, I will warn against the seduction of hypotheses that may ensure a positive answer to the third question, but may be hard to satisfy in specific models; see also the last paragraph in 5.4.B and remark 63(6) below.

DEFINITION 20. A state φ on the algebra $\mathcal{A}$ of quasi-local observables is said to be translation invariant whenever $\forall\,(a,A)\in G\times\mathcal{A}\,:\,\varphi(\nu_a[A])=\varphi(A)$, a situation denoted by $\varphi\circ\nu=\varphi$. The state φ is said to be extremal translation invariant if it is translation invariant and may not be written as a convex sum of different translation invariant states.

With G denoting the group of translations of $\mathbb{X}^n = \mathbb{R}^n$ or $\mathbb{Z}^n$, G is trivially identified with $\mathbb{X}^n$. Let $\mathcal{C}=\mathcal{C}(G)$ be the set of all complex-valued, continuous, bounded functions $f:G\to\mathbb{C}$. Henceforth, this set is equipped with the usual point-wise addition and multiplication of functions, and with the sup-norm $\|f\|=\sup_{x\in G}|f(x)|$. These operations equip $\mathcal{C}$ with the structure of an (abelian) C^*-algebra. Define then an action of G on $\mathcal{C}$ by $a[f](x)=f(x-a)$.

DEFINITION 21. With the above notations, an invariant mean on $\mathcal{C}$ is a state η on $\mathcal{C}$ such that $\forall\,(a,f)\in G\times\mathcal{C}\,:\,\eta(a[f])=\eta(f)$.

Given $\mathbb{X}^n$, there are several such means. For instance, the ergodic mean on $\mathbb{R}$ may be defined as follows. Let $\mathcal{C}_e=\{f\in\mathcal{C}\mid \lim_{a\to\infty}1/2a\int_{-a}^{a}dx\,f(x)\text{ exists}\}$. Then $\forall\,f\in\mathcal{C}_e$, let $\eta_e(f):=\lim_{a\to\infty}1/2a\int_{-a}^{a}dx\,f(x)$; which then extends by continuity to $\mathcal{C}$, so as to give an invariant mean, which is the one I will prefer to use in the sequel. One may wish to define similarly the mean η_+ on $\mathcal{C}_+=\{f\in\mathcal{C}\mid \lim_{x\to\infty}f(x)\text{ exists}\}$. And, similarly, another mean η_- obtains from the functions that admit a limit as $x\to-\infty$.

To define averages of states and of observables, notice that for every state φ on the algebra $\mathcal{A}$ of quasi-local observables and any $A,B\in\mathcal{A}$, the functions $\varphi(\nu_\bullet[A]B):a\in G\mapsto\varphi(\nu_a[A]B)\in\mathbb{C}$ — here the symbol $\bullet$ serves as a reminder to mark the place of the variable a — are continuous and bounded, namely by $\|A\|\,\|B\|$. Thus the functions $\varphi(\nu_\bullet[A]B)$ belong to $\mathcal{C}$. When $B=I$ we write simply $\varphi(\nu_\bullet[A])$ for $\varphi(\nu_\bullet[A]I)$. With these notations, the following definition makes sense.

DEFINITION 22. Given an invariant mean η on $\mathcal{C}$ and any state φ on the algebra $\mathcal{A}$ of quasi-local observables, the average $\eta[\varphi]$ of the state φ is defined as the translation invariant state

$$\eta[\varphi]:A\in\mathcal{A}\mapsto\eta\left(\varphi(\nu_\bullet[A])\right)\in\mathbb{C}\quad .$$

A translation invariant state φ is said to be $\eta-$clustering whenever

$$\forall\,A,B\in\mathcal{A}\,:\;\eta\left(\varphi(\nu_\bullet[A]\,B)\right)=\varphi(A)\,\varphi(B)\quad .$$

Warnings concerning terminology:

1. $\eta-$clustering is also referred to as “weak clustering”.

2. $\eta-$clustering should not be confused with the stronger property called “weak mixing”, namely

$$\forall\,A,B\in\mathcal{A}\,:\;\eta\;|\varphi(\nu_\bullet[A]\,B)\;-\;\varphi(A)\,\varphi(B)|=0$$

where for any complex number z , $|z|$ denotes absolute value of z . The name "weak mixing" conforms to the usage in classical ergodic theory, cf. e.g. [Arnold and Avez, 1968, p. 21].

3. The property simply called *clustering* does not involve averaging, and thus is stronger; it is:

$$\forall\, a \in \mathbb{R}^n \text{ and } \forall\, A, B \in \mathcal{A} \,:\, \lim_{\lambda \to \infty} \varphi(\nu_{\lambda a}[A]\, B) = \varphi(A)\, \varphi(B) \quad .$$

This property is called "mixing" in classical ergodic theory, cf. e.g. [Arnold and Avez, 1968, p. 20].

4. An even stronger property is introduced in definition 27 below.

5. Each of the above properties expresses how much the correlations between $\nu_a[A]$ and B decay with large distances $|a|$ when the system is in the state φ . The term "clustering" affixed to these properties, also used in QFT, seems to be inherited from scattering theory where it expresses the asymptotic independence of separate scattering products, or "clusters".

The definition of the average of an observable is a little bit more involved. For the general mathematical framework, cf. e.g. [Emch, 1972a, subsection 2.2.d]; in particular, for the general statements and proofs corresponding to scholium 23 and theorem 25 below, cf. [Emch, 1972a, lemma, pp. 174–175] and [Emch, 1972a, theorem 8, pp. 183–184]. Note that, here, the asymptotic abelianness of the action of the group of space translations — corollary 19 above — allows the simpler presentation offered below. *This is where global observables* — cf. definition 15 above — *enter the picture.*

Let φ be a translation invariant state on the algebra $\mathcal{A}$ of quasi–local observables, and $\{\pi_\varphi, \mathcal{H}, \Phi\}$ be the GNS triple associated to φ . Let further $\mathcal{N}_\varphi = \pi_\varphi(\mathcal{A})''$ and $\mathcal{Z}_\varphi = \mathcal{N}_\varphi \cap \mathcal{N}'_\varphi$.

For $a \in G$ fixed, and A running over $\mathcal{A}$, the map $\pi_\varphi(A)\Phi \in \mathcal{H} \mapsto \pi_\varphi(\nu_a[A])\Phi \in \mathcal{H}$ extends uniquely to a unitary operator $U_a \in \mathcal{U}(\mathcal{H}) := \{U \in \mathcal{B}(\mathcal{H}) \mid U^*U = UU^* = I\}$. This defines a continuous unitary representation $U : a \in G \mapsto U_a \in \mathcal{U}(\mathcal{H})$ such that $\forall\,(a, A) \in G \times \mathcal{A} \,:\, U_a \pi_\varphi(A) U_a^* = \pi_\varphi(\nu_a[A])$.

As usual, let $U(G)' := \{B \in \mathcal{B}(\mathcal{H}) \mid \forall\, a \in G \,:\, U_a B = B U_a\}$ denote the commutant of $U(G)$. Equivalently here, $U(G)' = \{B \in \mathcal{B}(\mathcal{H}) \mid \forall\, a \in G \,:\, U_a B U_a^* = B\}$.

Finally, let $\mathcal{P} := \{\Psi \in \mathcal{H} \mid \forall\, a \in G \,:\, U_a \Psi = \Psi\}$; and denote by P the orthogonal projector from $\mathcal{H}$ onto $\mathcal{P}$.

SCHOLIUM 23. For every invariant mean η on $\mathcal{C}$, the map

$$\eta_\varphi \,:\, A \in \mathcal{A} \mapsto \eta_\varphi[A] \in \mathcal{Z}_\varphi \cap U(G)'$$

defined, for all $A \in \mathcal{A}$ by

$$\forall\, \Psi_1, \Psi_2 \in \mathcal{H} \,:\, (\Psi_1, \eta_\varphi[A]\Psi_2) = \eta\,(\Psi_1, \pi_\varphi(\nu_\bullet[A])\Psi_2)$$

is a $^*-$homomorphism and satisfies $\eta_\varphi[A]\,P = P\,\eta_\varphi[A] = P\,\eta_\varphi[A]\,P$.

DEFINITION 24. Let η be an invariant mean on $\mathcal{C}$; φ be a translation invariant state on the algebra $\mathcal{A}$ of quasi-local obervables; $\mathcal{N}_\varphi = \pi_\varphi(\mathcal{A})''$ be the algebra of global observables associated to the state φ, via the GNS triple $\{\pi_\varphi, \mathcal{H}, \Phi\}$; and $\mathcal{N}_\varphi^G = \{N \in \mathcal{N}_\varphi \mid \forall\, a \in G \,:\, U_a N U_a^* = N\}$ be the algebra of translation invariant global observables. Then the average of a quasi-local observable $A \in \mathcal{A}$ is defined as the translation invariant global observable $\eta_\varphi[A] \in \mathcal{N}_\varphi^G$.

We are now ready to enunciate the central quantum ergodic theorem relative to the action of the group of space translations.

THEOREM 25. *Let* $\nu : a \in G \to \mathrm{Aut}(\mathcal{A})$ *denote the action of the space-translation group on the algebra* $\mathcal{A}$ *of quasi-local observables; and let* η *be any invariant mean on* $\mathcal{C}$ *. Then the following conditions on a translation invariant state* φ *on* $\mathcal{A}$ *are equivalent:*

1. φ *is extremal translation invariant;*
2. φ *is* $\eta-$*clustering, i.e.* $\forall\, A, B \in \mathcal{A} \,:\; \eta\,(\varphi(\nu_\bullet[A]\,B)) = \varphi(A)\,\varphi(B)$;
3. *the canonical extension* $\tilde{\varphi} : N \in \mathcal{N}_\varphi \mapsto (\Phi, N\,\Phi) \in \mathbb{C}$ *of* φ *to the von Neumann algebra* $\mathcal{N}_\varphi$ *of global observables associated to* φ *is the only translation invariant normal state on this algebra;*
4. *the invariant subspace* $\mathcal{P} \subset \mathcal{H}$ *is one-dimensional;*
5. *the average* $\eta_\varphi[A]$ *of every quasi-local observable* $A \in \mathcal{A}$ *is a multiple of the identity, namely* $\eta_\varphi[A] = \varphi(A)\,I$;
6. *all translation invariant global observables* $N \in \mathcal{N}_\varphi^G := \mathcal{N}_\varphi \cap U(G)'$ *are multiples of the identity;*
7. $\mathcal{Z}_\varphi \cap U(G)' = \mathbb{C}\,I$ *where* $\mathcal{Z}_\varphi := \mathcal{N}_\varphi \cap \mathcal{N}_\varphi{}'$.

REMARKS 26.

1. Recall that in definition 15 three kinds of observables were introduced. The *local* observables relative to some finite region Λ are described in the original von Neumann formalism [von Neumann, 1932c] where, typically, $\mathcal{A}_\Lambda = \mathcal{B}(\mathcal{H}_\Lambda)$, and $\mathcal{H}_\Lambda = \mathcal{L}^2(\Lambda, dx)$. Thus one refers to local observables as self-adjoint elements of $\mathcal{A}_o = \cup_{\Lambda \in \mathcal{F}} \mathcal{A}_\Lambda$. The *quasi-local observables,* defined abstractly as observables that are norm-limits of local observables, pertain to the microscopic description of many-body systems that are infinitely extended in space; section 5 below opens with three concrete QSP examples. These 'quasi-local' observables belong to the C^*-algebra $\mathcal{A} = {}^n\overline{\mathcal{A}_o}$. Observables of the third kind, the *global observables,* appear at the macroscopic level when bulk properties of matter are investigated; they belong to the von Neumann algebra $\mathcal{N}_\varphi := \pi_\varphi(\mathcal{A})''$ obtained as the weak-closure of the

GNS representation π_φ (of $\mathcal{A}$) corresponding to a state φ (on $\mathcal{A}$) specifically obtained by a process called the thermodynamical limit, several examples of which are discussed in the following sections.

Space averages are examples of such *global observables.* A concrete example in ferromagnetism obtains with any one of the three components of the magnetization. Observables of this third kind depend on the global state of the system considered, thus reflecting the preparation of the system. For instance, when the state is extremal translation invariant, these observables are multiples of the identity operator — recall the equivalence of conditions (1) and (5) in theorem 25 — hence their value is the same in all configurations that differ only locally from the given state. Their assuming different values in configurations that differ globally from one another serves as witness for the existence of different thermodynamical phases; cf. subsection 5.7.

2. A global state φ on $\mathcal{A} = {}^{n}\overline{\mathcal{A}_o}$ with $\mathcal{A}_o = \bigcup_{\Lambda\in\mathcal{F}} \mathcal{A}_\Lambda$ is usually defined by continuity from

$$(59)\quad \forall\,\Lambda \text{ and } \forall\, A_\Lambda \in \mathcal{A}_\Lambda \quad : \quad \varphi(A_\Lambda) = \lim_{\substack{|\Omega|\to\infty,\\ \Omega\in\mathcal{F},\Omega\supseteq\Lambda}} \varphi_\Omega(A_\Lambda)$$

where $\{\varphi_\Omega | \Omega \in \mathcal{F}\}$ is a consistent family of local states. The local states are themselves defined with respect to some consistent boundary conditions; e.g. periodic boundary conditions on every Λ. Hence, the global state φ and thus the von Neumann algebra $\mathcal{N}_\varphi := \pi_\varphi(\mathcal{A})''$ of global observables may depend on the boundary conditions one has chosen. This happens in particular in the presence of the long-range order that often accompanies the onset of phase transitions. This dependence on initial conditions, *even in the thermodynamical limit,* is an ubiquitous phenomenon, known already in classical statistical physics.

Indeed, in an argument that was later confirmed to be correct — for references, cf. e.g. [Emch and Liu, 2002, pp. 416–417] — Peierls [1936] pointed out the fact that the Ising model in two dimensions develops, for sufficiently low temperatures, a sensitivity to boundary conditions: one phase — say the one with strictly positive magnetization — may be selected by clamping all spins on the boundary in the "up" position.

3. Here again, in the special case where $\{\mathcal{N}, \varphi\}$ is $\{\mathcal{L}^\infty(\Omega), \mu\}$, the above theorem reduces to the known classical case. Note however that the theorem is stated here for *space translations* rather than for the *time evolution;* the reason is that the proof uses asymptotic abelianness which space translations satisfy — see corollary 19 above — or some weakened form such as (61) in remark 31 below. Yet, even such a weakened form of asymptotic abelianness is hard to come by for the time evolution of quantum dynamical models.

The clustering condition (2) in the theorem may be strengthened when the representation π_φ is primary, i.e. when the center $\mathcal{Z}_\varphi := \pi_\varphi(\mathcal{A})'' \cap \pi_\varphi(\mathcal{A})'$ satisfies $\mathcal{Z}_\varphi = \mathbb{C}I$. Specifically, for any region $\Lambda \in \mathcal{F}$, let

$$\mathcal{A}_\Lambda{}^c := {}^n\overline{\cup_{\Omega\in\mathcal{F};\Omega\bowtie\Lambda}\mathcal{A}_\Omega} ,$$

where, for any subset $\mathcal{B} \subset \mathcal{A}$, ${}^n\overline{\mathcal{B}}$ denotes the closure of $\mathcal{B}$ in the norm-topology of $\mathcal{A}$. As a consequence of locality $A \in \mathcal{A}_\Lambda$ and $B \in \mathcal{A}_\Lambda{}^c$ entail $AB - BA = 0$. Let now $\mathcal{N}_{\varphi,\Lambda}{}^c := \pi_\varphi(\mathcal{A}_\Lambda{}^c)''$.

DEFINITION 27. A state φ on the algebra $\mathcal{A}$ of quasi-local observables is said to be uniformly clustering whenever for any $A \in \mathcal{A}$ and every $\epsilon > 0$, there exists a region $\Lambda \in \mathcal{F}$ depending on A and ϵ, such that

(60) $\forall\, B \in \mathcal{A}_\Lambda{}^c \,:\, |\varphi(AB) - \varphi(A)\varphi(B)| \leq \epsilon\, \|B\|$.

DEFINITION 28. The elements of the von Neumann algebra $\mathcal{N}_\varphi{}^\infty := \bigcap_{\Lambda\in\mathcal{F}} \mathcal{N}_{\varphi,\Lambda}{}^c$ are called observables at infinity with respect to φ.

SCHOLIUM 29. For each state φ separately, the observables at infinity are central, i.e. $\mathcal{N}_\varphi{}^\infty \subseteq \mathcal{Z}_\varphi$. Moreover the following two conditions on a state φ are equivalent:

1. all observables at infinity are multiples of the identity operator, i.e. $\mathcal{N}_\varphi{}^\infty = \mathbb{C}\, I$;
2. φ is uniformly clustering.

Note that definitions 27, 28 and scholium 29 do not require that φ be space-translation invariant, although they involve in an essential manner the local structure of $\mathcal{A}$. For space-translation invariant states one has in addition:

COROLLARY 30. *The following two conditions:*

1. *φ is a translation invariant state on the algebra $\mathcal{A}$ of quasi-local observables;*
2. *the algebra $\mathcal{N}_\varphi$ of global observables is a factor, i.e. $\mathcal{Z}_\varphi = \mathbb{C}\, I$*

jointly entail that

a. *φ is extremal translation invariant (and so satisfies the equivalent conditions noted in theorem 25);*

b. *φ is uniformly clustering.*

REMARKS 31.

1. Condition (2) in corollary 30 is satisfied whenever φ is an extremal KMS state; cf. subsection 5.6 below.
2. The proofs of theorem 25, scholium 29, and corollary 30 are not trivial, but they were all known by the early 1970s; cf. e.g. [Emch, 1972a, theorem II.2.8 and theorem IV.1.7].

3. In particular, the proof of theorem 25 shows that the equivalence of its seven conditions may be obtained in more general contexts where the action of the group of space translations is replaced by an action with respect to which the invariant state φ satisfies the condition of $\eta-$*abelianness,* namely the condition:

   (61) $\forall\, A, B, C \in \mathcal{A} \,:\, \eta\,\{\varphi(C^*[\nu_g[A]\,B - B\nu_g[A]]C)\} = 0 \quad .$

   This condition is much weaker that the norm-asymptotic abelianness proven in corollary 19 for the action of the translation group.

4. It is therefore tempting to try and transfer the above considerations to the group $\mathbb{R}$ governing the time evolution of a quantum dynamical system. In fact *if* φ is an extremal $\mathbb{R}-$invariant state, then such a dynamical system will be $\eta-$abelian in the sense of (61), *provided* the vector Φ of the GNS representation — which, by construction, is cyclic for $\pi_\varphi(\mathcal{A})$ — is also cyclic for the von Neumann algebra $\pi_\varphi(\mathcal{A})'$, a condition equivalent to the requirement that Φ be separating for the von Neumann algebra $\mathcal{N}_\varphi := \pi_\varphi(\mathcal{A})''$, i.e. $N \in \mathcal{N}_\varphi$ and $N\Phi = 0$ entail $N = 0$. The condition that a von Neumann algebra $\mathcal{N}$ admits a vector Φ that is cyclic for both $\mathcal{N}$ and $\mathcal{N}'$ is referred to by saying that this von Neumann algebra is in standard form; for the relevance of this condition in the present context cf. definition 36 and theorem 39 below. This however only raises again the question of whether φ is extremal under the evolution responsible for the approach to equilibrium. In this respect, we may note that this is the case for the dilated evolution in the example of a chain of weakly coupled harmonic oscillators, discussed at the beginning of this subsection, and in general for the evolution α of generalized Kolmogorov flows; cf. definition 17; see nevertheless the "warning" following remark 18, or subsection 5.4(B).

5. On the mathematical side, quantum ergodic theory may be concerned with group actions more general than space or time translations. In fact, theorem 25 and the third remark just above extend without modifications to the actions of *amenable* groups, i.e. groups $\mathcal{G}$ that admit an invariant mean in the sense of definition 21 (where $G = \mathbb{R}^n$ or $\mathbb{Z}^n$ is replaced by $\mathcal{G}$). For a general presentation of the theory of amenable groups, cf. e.g. [Greenleaf, 1969] or for a brief review geared to applications in QSP [Emch, 1972a, pp. 164–172]. Restricting attention here to locally compact groups, let it suffice to note that compact groups, abelian groups, and semi-direct products thereof are amenable; in particular the rotation groups, translation groups, and Euclidean groups in finite-dimensional Euclidean spaces are amenable. However, *no* non-compact semi-simple Lie group is amenable, so that in particular the Lorentz group of 4-dimensional relativisitic QFT is *not* amenable.

6. Pushing the theory even further than amenable group actions may be done by considering "large groups of automorphisms" of a C^*-algebra $\mathcal{A}$, i.e.

actions $\alpha : \mathcal{G} \to \mathrm{Aut}(\mathcal{A})$ that satisfy for every self-adjoint $A \in \mathcal{A}$ and every $\mathcal{G}-$invariant state φ on $\mathcal{A}$:

$$(62) \quad {}^{w-op}\overline{co\{\pi_\varphi(\alpha_g[A]) \mid g \in \mathcal{G}\}} \cap \pi_\varphi(\mathcal{A})' \neq \emptyset \quad ,$$

where for any subset S of a vector space, $co\{S\}$ denotes the "convex hull" of S, i.e. the collection of all convex combinations of elements in S; and for any set $\mathcal{B} \subset \mathcal{B}(\mathcal{H})$, ${}^{w-op}\overline{\mathcal{B}}$ denotes the closure of $\mathcal{B}$ in the weak-operator topology of $\mathcal{B}(\mathcal{H})$. The notion of *large group of automorphisms* was introduced by Størmer in 1967 who used it soon afterwards to prove a quantum analogue of de Finetti's exchangeability theorem in classical probability theory [Størmer, 1969]; for a review and some applications to the semantic foundations of quantum theory, cf. e.g. [Emch, 2005] and references therein. Note that any amenable group action for which the system is $\eta-$abelian for some mean η is a large group of automorphisms for this system.

Here again, one can hardly resist the conclusion that quantum ergodic theory is now a mature mathematical theory in search of further physical applications to QSP, most notably through the understanding it provides for the various clustering (or mixing) properties described in the present section; cf. e.g. subsections 5.4 and 5.7 below.

4 THE KMS CONDITION FOR EQUILIBRIUM

The identification of the KMS condition as a canonical characterization of equilibrium states appears in the confluence of two currents of thought.

The first source is the recognition by Kubo [1957] and by Martin & Schwinger [1959] that objects which play a central role in condensed matter physics, namely the so-called thermal Green functions — cf. e.g. [Bonch-Bruevich and Tyablikov, 1962] — possess remarkable analytic properties. For a foretaste, see scholium 32 below.

The second source of inspiration is recognizable in the original texts [Murray and von Neumann, 1936] of what was to become the theory of von Neumann algebras, and is emphasized in the candid reminiscences of one of the pioneers of this theory [Murray, 1990]. A great deal of the theory could be built from the following observation: there are matrix algebras $\mathcal{N}$ which, together with their commutant $\mathcal{N}'$, satisfy the following properties:

(i) they are factors, i.e. have trivial center: $\mathcal{N} \cap \mathcal{N}' = \mathbb{C}I$; (ii) $\mathcal{N}$ and $\mathcal{N}'$ admit a common cyclic vector Φ; (iii) there exists an involutive antiunitary operator J such that $J\Phi = \Phi$ and $N \in \mathcal{N} \mapsto JNJ \in \mathcal{N}'$ is bijective. For a concrete, simple example, see equation (71) below.

Each of the two facets of the theory — analytic and algebraic — involves some mathematical intricacies; hence the division of this section into two subsections: first, a simple example; and second, the general theory.

4.1 A Wignerian Approach

In this subsection, I wish to abide by Wigner's famous dictum [Wigner, 1962]: *"Please explain it with* 2×2 *matrices."* Accordingly I proceed with the description of what happens to a quantum $1/2-$spin in canonical equilibrium at natural temperature $\beta > 0$ in a magnetic field B parallel to the z-axis. The observables are the self-adjoint elements of the algebra $\mathcal{M}$ of 2×2 matrices with complex entries. The Hamiltonian is

$$(63) \quad H = -B\sigma^z = \begin{pmatrix} \epsilon_1 & 0 \\ 0 & \epsilon_2 \end{pmatrix} \quad \text{with} \quad \epsilon_1 = -B \ , \ \epsilon_2 = +B.$$

The canonical equilibrium state is, according to von Neumann's characterization (38):

$$(64) \quad \psi_H : M \in \mathcal{M} \to \mathrm{Tr}(\rho_H M) \quad \text{with} \quad \rho_H = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$$

where $\lambda_n = {Z_H}^{-1} \exp(-\beta\epsilon_n)$, with $Z_H = \exp(-\beta\epsilon_1) + \exp(-\beta\epsilon_2)$ denoting the partition function of the system.

In the Heisenberg picture, conjugate to Schrödinger's picture (39), the evolution is

$$(65) \quad \left. \begin{array}{l} \alpha_t : M \in \mathcal{M} \to \alpha_t[M] = U^*(t) M U^*(-t) \in \mathcal{M} \\ \\ \qquad \text{with} \quad U^*(t) = \begin{pmatrix} e^{i\epsilon_1 t} & 0 \\ 0 & e^{i\beta\epsilon_2 t} \end{pmatrix} \ . \end{array} \right\}$$

To make computations easier and, moreover, immediately generalizable to higher dimensions, consider the matrices

$$E_{mn} : \Psi \in \mathbb{C}^2 \mapsto (\Psi_n, \Psi)\Psi_m \in \mathbb{C}^2$$

where $\{\Psi_n \mid n = 1,2\}$ are eigenvectors of H , i.e. with

$$\Psi_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \ ; \ \Psi_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \ :$$

$$E_{11} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}; \quad E_{12} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}; \quad E_{21} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}; \quad E_{22} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

These matrices form a basis in $\mathcal{M}$ and — with ψ_H and α_t as in (64) and (65) — satisfy

$$E_{kl}E_{mn} = \delta_{lm}E_{kn} \ , \ \psi_H(E_{mn}) = \lambda_m \delta_{mn} \ , \ \alpha_t(E_{mn}) = e^{i(\epsilon_m - \epsilon_n)t} E_{mn} \ .$$

From these relations and the identity $\exp[-\beta(\epsilon_m - \epsilon_n)]\lambda_n = \lambda_m$, one obtains that the analytic functions

$$f_{klmn} : z \in \mathbb{C} \to \lambda_n \, e^{i(\epsilon_m - \epsilon_n)z} \delta_{lm}\delta_{kn}$$

satisfy $\forall\, t \in \mathbb{R}: \; f_{klmn}(t) = \psi_H(E_{kl}\alpha_t[E_{mn}])$ and $f_{klmn}(t{+}i\beta) = \psi_H(\alpha_t[E_{mn}]E_{kl})$. Moreover, on the strip

$$\Omega_\beta := \{z \in \mathbb{C} \mid 0 \leq \operatorname{Im} z \leq \beta\}\,,$$

the analytic functions f_{klmn} are bounded, namely by $\exp(|\epsilon_m - \epsilon_n|\,\beta)$.

These two properties of the canonical equilibrium state ψ_H extend by linearity to the time correlation functions

(66) $f_{MN}(t) = \psi_H(M\alpha_t[N])$ and $f_{MN}(t + i\beta) = \psi_H(\alpha_t[N]M)$

with M and N arbitrary in $\mathcal{M}$.

Conversely, suppose that φ is a state on $\mathcal{M}$ such that for every pair M, N of elements in $\mathcal{M}$ there exists a function $f_{M,N} : z \in \Omega_\beta \mapsto f_{M,N}(z) \in \mathbb{C}$ such that

(i) $f_{M,N}$ is bounded and continuous on the strip Ω_β ;

(ii) $f_{M,N}$ is analytic inside that strip;

(iii) for all $t \in \mathbb{R} : f_{M,N}(t) = \varphi(M\alpha_t[N])$ and $f_{M,N}(t + i\beta) = \varphi(\alpha_t[N]M)$.

Then in particular, with $M = I$, the function $f_{I,N}$ is periodic with period $i\beta$. It may then be extended to a function that is both bounded and analytic on the whole complex plane. The classical Liouville's theorem — cf. e.g. [Churchill and Brown, 1990, theorem 43.1] — thus entails that this function must be constant, i.e. for all $(t, N) \in \mathbb{R} \times \mathcal{M} : \varphi(\alpha_t[N]) := \operatorname{Tr} U^*(-t)\rho U^*(t)N$ is equal to $\operatorname{Tr}\rho N = \varphi(N)$; and thus

$$\rho = \begin{pmatrix} \mu_1 & 0 \\ 0 & \mu_2 \end{pmatrix}$$

where the values of μ_1 , μ_2 positive with $\mu_1 + \mu_2 = 1$ are computed presently. Upon comparing, for every pair of indices (m, n) the analytic continuation of

$$f_{mn}(t) = \varphi(E_{nm}\alpha_t[E_{mn}]) = e^{i(\epsilon_m - \epsilon_n)t}\mu_n$$

and

$$f_{mn}(t + i\beta) = \varphi(\alpha_t[E_{mn}]E_{nm}) = e^{i(\epsilon_m - \epsilon_n)t}\mu_m$$

one obtains $\exp[-\beta(\epsilon_m - \epsilon_n)]\mu_m = \mu_n$ and thus, upon imposing the normalization $\varphi(I) = 1$, i.e. $\mu_1 + \mu_2 = 1$:

$$\mu_n = \frac{e^{-\beta\mu_n}}{e^{-\beta\mu_1} + e^{-\beta\mu_2}} = \lambda_n \quad .$$

Thus, indeed $\varphi = \psi_H$.

In summary, one obtained by elementary means an elementary illustration of the first facet of the theory, its analytic aspect:

SCHOLIUM 32. Let $H = -B\sigma^z$ be the Hamiltonian describing a spin $\frac{1}{2}$ in a magnetic field B . Then, for any state φ on $\mathcal{M} = \mathcal{M}(2, \mathbb{C})$, the following conditions are equivalent:

(I) φ is the canonical equilibrium state ψ_H with respect to the Hamiltonian H ;

(II) for every pair (M, N) of elements of $\mathcal{M}$ there exists a function $f_{M,N} : z \in \Omega_\beta \to \mathbb{C}$ such that

$$\text{(67)} \quad \left.\begin{array}{l} f_{M,N} \textit{ is bounded and continuous on } \Omega_\beta \,; \\ f_{M,N} \textit{ is analytic in the interior of } \Omega_\beta \,; \\ \forall\, t \in \mathbb{R} : \left\{ \begin{array}{ll} f_{M,N}(t) & = \varphi(M\,\alpha_t[N]) \\ f_{M,N}(t+i\beta) & = \varphi(\alpha_t[N]\,M) \end{array}\right. \end{array}\right\} \quad .$$

Moving now towards the algebraic aspect of the theory, one pursues with the same simple model, and let φ be a faithful state over $\mathcal{M}$, i.e. a state such that $M \in \mathcal{M}$ and $\varphi(M^* M) = 0$ entail $M = 0$. Without loss of generality one may choose a basis in which the density matrix ρ corresponding to φ is diagonal, with eigenvalues λ_n $(n = 1, 2)$ strictly positive since φ is supposed to be faithful. Consider the representation π of $\mathcal{M}$ given by:

$$\text{(68)} \quad \forall\, M = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathcal{M} \quad : \quad \pi(M) = \begin{pmatrix} a & b & 0 & 0 \\ c & d & 0 & 0 \\ 0 & 0 & a & b \\ 0 & 0 & c & d \end{pmatrix} = M \otimes I$$

acting on the Hilbert space $\mathbb{C}^4$ equipped with its standard scalar product in which Ψ_{kl} defined by

$$\Psi_{11} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \quad , \quad \Psi_{21} = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} \quad , \quad \Psi_{12} = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} \quad , \quad \Psi_{22} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}$$

is an orthonormal basis. The vector

$$\text{(69)} \quad \Phi = \sum_k {\lambda_k}^{\frac{1}{2}} \Psi_{kk} = \begin{pmatrix} {\lambda_1}^{\frac{1}{2}} \\ 0 \\ 0 \\ {\lambda_2}^{\frac{1}{2}} \end{pmatrix}$$

satisfies $\Psi_{mn} = {\lambda_n}^{\frac{1}{2}} \pi(E_{mn})\Phi$, from which one reads:

$$\mathbb{C}^4 = \{\pi(M)\Phi \mid M \in \mathcal{M}\} \quad \text{and} \quad \forall\, M \in \mathcal{M} : (\Phi, \pi(M)\Phi) = \varphi(M) \quad .$$

Hence $\{\mathcal{H} := \mathbb{C}^4, \pi, \Phi\}$ is the canonical GNS triple associated to the state φ. Moreover, since φ is assumed to be faithful, $\|\pi(M)\Phi\| = 0$ entails $M = 0$, i.e. Φ is also separating for $\pi(\mathcal{M})$. The essential step now is to introduce the two operators J and Δ defined on $\mathcal{H}$ by the conditions that J is antilinear, Δ is linear, with

$$J\Psi_{mn} = \Psi_{nm} \quad \text{and} \quad \Delta\Psi_{mn} = \frac{\lambda_m}{\lambda_n}\Psi_{mn} \ .$$

Note that, since Δ is given here with its spectral resolution, the functions of this operator may be defined by linearity from $f(\Delta) : \Psi_{mn} \in \mathbb{C}^4 \mapsto f(\frac{\lambda_m}{\lambda_n})\Psi_{mn} \in \mathbb{C}^4$. In particular, $\{\Delta^{is} | s \in \mathbb{R}\}$ is a continuous group of unitary operators acting on $\mathbb{C}^4$.

One verifies immediately from their definition above that the operators J and Δ satisfy the following properties. Firstly,

(70) J is an isometry, $J^2 = I$, Δ is self-adjoint, $J\Delta J = \Delta^{-1}$, $J\Phi = \Phi = \Delta\Phi$.

Secondly,

(71) $$J \begin{pmatrix} a & b & 0 & 0 \\ c & d & 0 & 0 \\ 0 & 0 & a & b \\ 0 & 0 & c & d \end{pmatrix} J = \begin{pmatrix} a^* & 0 & b^* & 0 \\ 0 & a^* & 0 & b^* \\ c^* & 0 & d^* & 0 \\ 0 & c^* & 0 & d^* \end{pmatrix} \in I \otimes \mathcal{M}$$

i.e. upon denoting by $\mathcal{N}$ the image $\pi(\mathcal{M})$ of $\mathcal{M}$ through the representation π, we have: $J \mathcal{N} J = \mathcal{N}'$; hence (71) gives an explicit bijection from $\mathcal{N}$ onto its commutant $\mathcal{N}'$. The relation (71) is a particular case of the general Tomita–Takesaki duality (see theorem 39 below).

Thirdly, with $\beta > 0$ arbitrary, but fixed, we have $\forall t \in \mathbb{R} : \Delta^{-it/\beta}\Psi_{mn} = \exp[i(\epsilon_m - \epsilon_n)t]\,\Psi_{mn}$. Hence $\Delta^{-it/\beta}\pi(E_{mn})\Delta^{it/\beta}\Psi_{kl} = \exp[i(\epsilon_m - \epsilon_n)t]\pi(E_{mn})\Psi_{kl}$ with $\epsilon_n = c - (1/\beta)\ln\lambda_n$ where c is an arbitrary real constant. Consequently, the unitary group $\{\Delta^{it/\beta} \mid t \in \mathbb{R}\}$ implements a group of automorphisms of $\mathcal{N}$, namely

(72) $\tau_t : N \in \mathcal{N} \to \tau_t[N] = \Delta^{-it/\beta} N \Delta^{it/\beta} \in \mathcal{N}$

with, for all $(t, M) \in \mathbb{R} \times \mathcal{M}$, $\tau_t[\pi(M)] = \pi(\alpha_t[M])$ with $\alpha_t[M] = \exp^{iHt} M \exp^{-iHt}$ and $H = \sum_n \epsilon_n E_{nn}$. Summing up, this establishes that φ is the canonical equilibrium state at natural temperature β for the Hamiltonian H just constructed.

Fourthly, the operator $S = J\Delta^{\frac{1}{2}}$ satisfies $S\,\pi(E_{mn})\Phi = \pi(E_{nm})\Phi$ and thus, since J and therefore S are antilinear:

(73) $\forall N \in \mathcal{N} : \quad S\,N\,\Phi = N^*\,\Phi \quad .$

Finally, the generator L of the unitary group $\{\Delta^{it/\beta} \mid t \in T\}$ on $\mathbb{C}^4 = \mathbb{C}^2 \otimes \mathbb{C}^2$ is

(74) $L = H \otimes I - I \otimes H$,

so that the spectrum of L is symmetric around 0 : $Sp(L) = \{\epsilon_2 - \epsilon_1, 0, 0, \epsilon_1 - \epsilon_2\}$.

SCHOLIUM 33. Let $\{\mathbb{C}^4, \pi, \Phi\}$ be the GNS triple canonically associated to a faithful state φ on the algebra $\mathcal{M}$ of 2×2 matrices; and let $\mathcal{N}$ be the von Neumann algebra $\pi(\mathcal{M}) = \{\pi(M) \mid M \in \mathcal{M}\}$ acting on $\mathcal{H} = \mathbb{C}^4$. Then

1. $\mathcal{N}$ is isomorphic to $\mathcal{M}$ and φ may be regarded as a faithful state on $\mathcal{N}$;

2. Φ is both cyclic and separating for $\mathcal{N}$;

3. the anti-linear operator defined by $S : N\Phi \in \mathcal{H} = N^*\Phi \in \mathcal{H}$ has polar decomposition $S = J\Delta^{\frac{1}{2}}$ where J is an involutive, anti-linear isometry from $\mathcal{H}$ onto itself, and Δ is a positive operator acting on $\mathcal{H}$;

4. J establishes a duality between $\mathcal{N}$ and its commutant; specifically: $N \in \mathcal{N} \mapsto JNJ \in \mathcal{N}'$ is an anti-linear bijection;

5. $\{\Delta^{-it/\beta} \mid t \in \mathbb{R}\}$ implements a group of automorphisms τ_t of $\mathcal{N}$ with respect to which the state φ satisfies the analyticity condition described in scholium 32;

6. $J\Phi = \Phi = \Delta\Phi \quad , \quad J^2 = J \quad \text{and} \quad \forall\, s \in \mathbb{R} : J\Delta^{is} = \Delta^{is}J \quad .$

REMARKS 34. Upon surveying the proofs of scholia 32 and 33, one verifies that they can be extended verbatim from $\mathcal{M} = \mathcal{M}(2,\mathbb{C})$ to $\mathcal{M} = \mathcal{M}(n,\mathbb{C})$ where n is any finite positive integer. These scholia extend further to $\mathcal{M} = \mathcal{B}(\mathcal{H})$ where $\mathcal{H}$ is a Hilbert space (with countable basis) provided that:

(i) the Hamiltonian H satisfies $\mathrm{Tr}(-\beta H) < \infty$;

(ii) the state φ is countably additive, retaining the condition that φ be faithful.

Indeed, under these circumstances one can read again the proofs of the scholia, now for the *−algebra $\mathcal{E} = \mathrm{Span}\{\pi(E_{mn}) \mid m, n = 1, \ldots\}$ of all finite linear combinations of the operators $\pi(E_{mn})$ with $E_{mn} : \Psi \in \mathcal{H} \to (\Psi_n, \Psi)\Psi_m \in \mathcal{H}$ where again $\{\Psi_n \mid n = 1, 2, \ldots\}$ is an orthonormal basis in $\mathcal{H}$. The extension from $\mathcal{E}$ to the von Neumann algebra $\mathcal{B}(\mathcal{H})$ obtains since the assumptions that φ is countably additive and faithful allow one to use standard continuity arguments, namely here, e.g. [Dixmier, 1957, theorem I.3.5, lemma I.4.4, proposition I.4.1]; or [Kadison and Ringrose, 1983/1986, volume ii, chapter 7]. In particular, $\mathcal{N} = \pi(\mathcal{B}(\mathcal{H})) = \{\pi(\mathcal{M}) \mid M \in \mathcal{B}(\mathcal{H})\}$ is already a von Neumann algebra — i.e. $\mathcal{N} = \mathcal{N}''$ — and is isomorphic to $\mathcal{B}(\mathcal{H})$. Since $\mathcal{B}(\mathcal{H})$ is a factor, so is $\mathcal{N}$, i.e. the center of this von Neumann algebra is trivial: $\mathcal{N} \cap \mathcal{N}' = \mathbb{C}I$. Moreover $\mathcal{N}$ may be identified with $\mathcal{B}(\mathcal{H}) \otimes \mathbb{C}I$ and $\mathcal{N}'$ with $\mathbb{C}I \otimes \mathcal{B}(\mathcal{H})$.

The von Neumann formalism for quantum mechanics [von Neumann, 1932c] allows one to go this far, but no further. Recall that some of the reasons why one needs to proceed further were indicated in subsection 3.3. The next subsection provides an important tool toward achieving this.

4.2 The Kubo–Martin–Schwinger condition and the Tomita–Takesaki theory

The above results suggest three definitions; the first two are just matters of mathematical terminology, but the third is at the heart of this section.

DEFINITION 35. A state φ on a von Neumann algebra $\mathcal{N}$ is said to be normal whenever it is countably additive., i.e. $\varphi(\sum_n P_n) = \sum_n \varphi(P_n)$ for each countable family $\{P_n\}$ of mutually orthogonal projections in $\mathcal{N}$.

This simply extends to general von Neumann algebras condition (35), already recognized in [von Neumann, 1932c] as the quantum analogue of the complete additivity of probability measures. The next definition formalizes in the present context some of the notions encountered in the motivating examples covered in the previous subsection.

DEFINITION 36. A von Neumann algebra $\mathcal{N}$ acting on a Hilbert space $\mathcal{H}$ is said to be in standard form whenever there exists a vector $\Phi \in \mathcal{H}$ that is both cyclic and separating for $\mathcal{N}$, i.e. $\mathcal{N}\Phi$ is norm dense in $\mathcal{H}$ and for $N \in \mathcal{N}$, $N\Phi = 0$ entails $N = 0$.

REMARKS 37. This concept has been around for a long time, but it seems fair to say that full recognition of its central importance in the general theory of von Neumann algebras had to wait for the Tomita–Takesaki modular theory [Tomita, 1967; Takesaki, 1970a]. At the most basic level, notice that if $\mathcal{N}$ is in standard form, one may assume without loss of generality that $\|\Phi\| = 1$, so that $\varphi : N \in \mathcal{N} \to (\Phi, N\Phi) \in \mathbb{C}$ is a faithful normal state on $\mathcal{N}$.

Conversely it follows, from the same continuity arguments as those used in remark 34 above, that if φ is any normal state on a von Neumann algebra $\mathcal{N}$, the GNS representation π corresponding to φ is already a von Neumann algebra; if φ is faithful, then $\mathcal{N}$ is isomorphic to $\pi(\mathcal{N})$. Thus the canonical GNS vector Φ is not only cyclic, but it is also separating. Hence whenever φ is a faithful normal state, $\mathcal{N}$ is isomorphic to $\pi(\mathcal{N})$ which is a von Neumann algebra presented in standard form

The third definition pertains to the core of this section. It is an adaptation of the work of [Kubo, 1957; Martin and Schwinger, 1959], proposed by [Haag *et al.*, 1967] as an extension of the definition of canonical equilibrium states on the global C^*-algebra to be associated to an infinite system.

DEFINITION 38. Let $\mathcal{A}$ be a C^*-algebra, and let $\alpha : t \in \mathbb{R} \to \alpha_t \in \mathrm{Aut}(\mathcal{A})$ be a group of automorphisms of $\mathcal{A}$. A state φ on $\mathcal{A}$ is said to satisfy the KMS condition with respect to α for the natural temperature β if for every pair (A, B) of elements of $\mathcal{A}$ there exists a function $f_{A,B}$ defined on the strip $\Omega_\beta = \{z \in \mathbb{C} \mid 0 \leq \mathrm{Im} z \leq \beta\}$, such that $f_{A,B}$ is bounded and continuous on Ω_β; $f_{A,B}$ is analytic in the interior of Ω_β; and $\forall\, t \in \mathbb{R} : f_{A,B}(t) = \varphi(A\, \alpha_t[B])$ and $f_{A,B}(t + i\beta) = \varphi(\alpha_t[B]\, A)$.

The main mathematical result of this section, taken from the Tomita–Takesaki modular theory [Tomita, 1967; Takesaki, 1970a], may now be stated.

THEOREM 39 (Tomita–Takesaki). *Let $\mathcal{N}$ be a von Neumann algebra acting on a Hilbert space $\mathcal{H}$ and admitting a cyclic and separating unit vector Φ. Then the closed antilinear operator S obtained as the closure of the map $N\,\Phi \to N^*\,\Phi$, defined for all $N \in \mathcal{N}$, has polar decomposition $S = J\Delta$ where $J = J^2$ is an antilinear isometry from $\mathcal{H}$ onto itself, satisfying $J\mathcal{N}J = \mathcal{N}'$; and Δ is a self-*

adjoint operator (not necessarily bounded!) that is positive, and such that $J\Delta^{it} = \Delta^{it}J$ *; and for any* $\beta > 0$ $\forall (t, N) \in \mathbb{R} \times \mathcal{N} : \tau_t[N] = \Delta^{-it/\beta} N \Delta^{it/\beta}$ *defines a group* $\{\tau_t\}$ *of* $^*-$*automorphisms of* $\mathcal{N}$ *with respect to which the faithful normal state* $\varphi : N \in \mathcal{N} \to (\Phi, N\Phi) \in \mathbb{C}$ *satisfies the KMS condition for* β. *Moreover* $\{\tau_t \mid t \in \mathbb{R}\}$ *is the unique group of* $^*-$*automorphisms of* $\mathcal{N}$ *with respect to which* φ *satisfies this condition.*

REMARKS 40.

1. The theorem generalizes to any arbitrary von Neumann algebra in standard form the result we described — in remark 34 — for the GNS representation of $\mathcal{B}(\mathcal{H})$ associated to any of its faithful normal states.

2. It is essential to the purpose of the present review to emphasize that the theorem does *not* require that $\mathcal{N}$ be a factor.

3. Whereas the theorem asserts that the dynamics τ is uniquely determined by the KMS condition, the converse is *not* true: when $\mathcal{N}$ is not a factor, there exist other normal states on $\mathcal{N}$ that also satisfy the KMS condition with respect to the same dynamics. Indeed, when $\mathcal{N}$ is not a factor, one verifies that for every $Z \neq 0$ that belongs to the center $\mathcal{Z} = \mathcal{N} \cap \mathcal{N}'$, $\psi : N \in \mathcal{N} \to [\varphi(Z^*Z)]^{-1}\varphi(Z^*NZ)$ defines a normal state that again satisfies the KMS condition with respect to τ for the same β. This remark, the proof of which will be given in subsection 5.6, is essential to the arguments presented in subsection 5.7.

4. Beyond its mathematical attractiveness, the legitimacy of the conjecture that the KMS condition may be regarded as a definition of canonical equilibrium states in the QSP of macroscopic systems will also be discussed in the next section.

5. Finally, mathematical probity requires us to mention that — factor or not — a major difficulty in the proof of theorem 39 resides in showing that the map $N\Phi \to N^*\Phi$ is closable; for the resolution of this problem, cf. the original papers [Tomita, 1967; Takesaki, 1970a]; it is probably fair to warn the reader that even the didactic presentation in [Kadison and Ringrose, 1983/1986, chapter 9] would have carried us beyond the bounds of this essay. To convey nevertheless an idea of the structures involved in the theorem, I resorted therefore to presenting first the models covered in the preliminary scholia 32 and 33, as these could be treated with mathematically elementary tools. The drawback was however that these models, as well as their routine extensions from $\mathcal{M}(2,\mathbb{C})$ to $\mathcal{B}(\mathcal{H})$ described in remark 34, only involve factors, in fact faithful representations of $\mathcal{B}(\mathcal{H})$, that are not sufficient to cover the macroscopic purposes of QSP where infinitely many degrees of freedom are brought to play. As Haag, Hugenholtz, and Winnink [1967] correctly envisaged, it is the generality involved in theorem 39 that is actually needed in physical applications. The temporal coincidence of this physical intuition

and the arrival on the scene of the mathematical theory of Tomita–Takesaki [1967; 1970a] is a truly remarkable event vividly recounted in [Kadison, 1990, pp. 77–79].

5 KMS CONDITION, QSP AND THERMODYNAMICS

This section presents some of the evidences supporting the physical interpretation of the KMS condition proposed in [Haag *et al.*, 1967] as an alternative definition of equilibrium states in QSP. We already saw that for finite systems the KMS condition is satisfied by the canonical equilibrium states of von Neumann, and only by those states. Now, in subsections 5.1–5.3 models are described to show how the modular structures invented and developed by [Tomita, 1967; Takesaki, 1970a] — which we saw (cf. scholium 33) are realized in finite systems in canonical equilibrium — are also encountered in the equilibrium QSP of infinite systems, thus allowing one to go beyond von Neumann's formalism [von Neumann, 1932c]. In subsection 5.4 various stability conditions are exhibited that give a thermodynamical characterization of KMS states in QSP. A brief excursion is undertaken in subsection 5.5 to indicate some vistas toward the recognition of the role the KMS condition has later been called to play in relativistic QFT, a role dubbed "revolutionary" by the practitioners. Subsection 5.6 is a mathematical interlude devoted to the algebraic characterization of *extremal* KMS states. When we return to QSP in subsection 5.7, systems that exhibit phase transitions are considered and the unique decomposition of any canonical equilibrium state into its pure thermodynamical phases is shown to be closely modeled by a unique decomposition of KMS states into extremal KMS states. In particular, this subsection is oriented toward substantiating the overarching idea that the KMS condition provides the thermodynamics of infinite systems with a conceptual scheme in which phase transitions occur accompanied by spontaneous symmetry breakdown.

5.1 Beyond Fock space: The BCS model

The first indication that something was amiss in the use of the von Neumann formalism in QSP was the Bardeen–Cooper–Schrieffer model for superconductivity, the BCS model. Indeed, in the original treatment of this model [Bardeen *et al.*, 1957], the Hamiltonian chosen to describe a specific interaction between the electrons in a finite but large metallic solid is invariant under gauge transformations of the first kind; an approximation is then proposed, which is asserted to become exact in the infinite volume limit; in this formal process however this symmetry is lost; moreover, the spectrum of the resulting Hamiltonian presents an energy gap that is temperature–dependent. One might argue that the experimentalist may not wish to be concerned with the breaking of that symmetry, but the energy gap cannot be ignored: experimentalists do measure it in the laboratory. Thus, mathematical physicists thought that they ought to understand — how or rather whether — the Hamiltonian itself may indeed depend on the temperature. Within five years, the culprit was found by Haag [1962] to be that the whole treatment was allegedly carried out in a fixed irreducible representation of the CCR, the then ubiquitous Fock representation, and that this constraint was doing violence to the model.

Specifically, the original Hamiltonian is

$$(75)\quad H_\Lambda = \sum_{p,s} \epsilon(p) a_s(p)^* a_s(p) \; + \; \sum_{p,q} b(p)^* \tilde{v}(p,q) b(q)$$

where Λ is the region of space in which the system is contained, typically a cubic box of finite volume $|\Lambda|$; p and q label momentum and are integer multiples of $2\pi|\Lambda|^{-\frac{1}{2}}$; $s = \pm\frac{1}{2}$; $a_s(p)^*$ and $a_s(p)$ are the creation and annihilation operators for an electron of spin s and momentum p; $\epsilon(p) = -\mu + \frac{1}{2}p^2/2m$ is the energy of a free electron of momentum p; $b(p)^* = a_\uparrow(p)^* a_\downarrow(-p)^*$ is the creation operator of a so-called Cooper pair; and $b(p)^*\tilde{v}(p,q)b(q)$ is the interaction energy between two Cooper pairs, i.e. four electrons, so that the Hamiltonian (75) is *not* quadratic in the original field operators. The form of $\tilde{v}(p,q)$ will be discussed later on.

The approximating Hamiltonian is

$$(76)\quad \tilde{H}_\Lambda = \sum_{p,s} E(p) \gamma_s(p)^* \gamma_s(p)$$

where $\gamma_s(p)^*$ and $\gamma_s(p)$ are the creation and annihilation operators for the elementary excitations given by a Bogoliubov–Valatin transformation

$$(77)\quad \left.\begin{array}{ll} \gamma_\uparrow(p) & = u(p) a_\uparrow(p) + v(p) a_\downarrow(-p)^* \\ \gamma_\downarrow(p) & = -v(-p) a_\uparrow(-p)^* + u(-p) a_\downarrow(p) \end{array}\right\}$$

where

$$(78)\quad \begin{array}{ll} E(p) & = \{\epsilon(p)^2 + [\Delta(p)\Delta(p)^*]\}^{\frac{1}{2}} \\ D(p) & = \{[E(p)-\epsilon]^2 + [\Delta(p)\Delta(p)^*]\}^{\frac{1}{2}} \end{array} \quad ; \quad \begin{array}{ll} u(p) & = \Delta(p)^*/D(p) \\ v(p) & = [E(p)-\epsilon(p)]/D(p) \end{array}$$

and Δ satisfies the all-important self-consistency equation

$$(79)\quad \Delta(p) = -\sum_q \tilde{v}(p,q) \frac{\Delta(q)}{2E(q)} \tanh\left(\frac{1}{2}\beta E(q)\right) \quad .$$

Clearly $\Delta = 0$ is always a solution, in which case the spectra of H and $\tilde{H}$ coincide; this is the normal phase in which nothing particularly interesting happens. The essence of the model is that there is a critical temperature T_c (recall $\beta = 1/kT$) below which an energetically more favorable solution $\Delta \neq 0$ develops. This corresponds to the superconducting phase. We henceforth pursue the discussion for $0 < T < T_c$.

This is the phase we are interested in, and it may be useful to recall in physical terms what the physicists first saw in (76)–(79). BCS devised a limiting procedure — involving the thermodynamical limit and a "mean-field approximation" (weak, but very long range interaction) — by which the original Hamiltonian (75) and the new Hamiltonian (76) become interchangeable in the sense that they are claimed to lead to the same limit. While (75) is expressed in terms of the electrons' creation and annhilation operators $a_s^\flat(p)$; the new Hamiltonian (76) is free in terms of the elementary excitations $\gamma_s^\flat(p)$. The energy spectrum of these excitations is $\{E(p)\}$ and differs — see (78) — from the energy spectrum $\{\epsilon(p)\}$ of the free electrons by a temperature-dependent "gap" which is observable in the laboratory; the numerical results so obtained for this gap are in very good agreement with the prediction (79); cf. [Schrieffer, 1974, Figure 1–3].

The mathematical picture however demands some explanation. Indeed: (i) the initial Hamiltonian (75) is invariant under the gauge symmetry defined, for any $\theta \in (0, 2\pi]$ by $a_s(p) \to \exp(i\theta)a_s(p)$ whereas the Hamiltonian (76) is not; and (ii) the energy spectrum $\{E(p)\}$ of the Hamiltonian (76) is temperature dependent, whereas there is no temperature dependence in (75).

The question therefore is to account for how one could possibly claim — as was done in the prevailing folklore — that such an approximation could become exact in the thermodynamical limit. For this, one has to examine where Δ comes from, namely that $\Delta(p)$ is a scalar multiple of the identity operator, to be viewed as an approximation of the operator $\hat{\Delta}(p) = \sum_q \tilde{v}(p,q)b(q)$. The argument for this is based on the remark that, under suitable assumptions on $\tilde{v}$, one can arrange for the limit $|\Lambda| \to \infty$ of $\hat{\Delta}(p)$ to exist — in the weak-operator topology — and to commute with all the creation and annihilation operators $a_s(q)^*$ and $a_s(q)$ which generate an algebra which is tacitly assumed to be irreducible. In this limit, the operator $\Delta(p)$ would be replaced by a scalar multiple of the identity. Some "suitable" assumptions seemed to be achieved when $\tilde{v}$ is the double Fourier transform

$$\tilde{v}(p,q) = \int_\Lambda dxdy f(p,x)v(x,y)f(q,y)^* \quad \text{where} \quad f(p,x) = \begin{cases} |\Lambda|^{-\frac{1}{2}}\, e^{ipx} & x \in \Lambda \\ 0 & x \notin \Lambda \end{cases}$$

with a nonlocal potential v such that $v(x,y)^* = v(y,x)$, $c = \int dxdy|v(x,y)| < \infty$ and $\sum_q |\tilde{v}(p,q)| < \infty$, so that $\lim_{\Lambda\to\infty} |\tilde{v}(p,q)| = 0$ and $|\tilde{v}(p,q)| \leq c/|\Lambda|$.

The practitioner will recognize here an approximation of the mean molecular field type, a heuristic tool introduced, during the first ten years of the twentieth century, by P. Weiss and L.S. Ornstein in the classical theory of phase transitions. Yet, the approximation is not acceptable here without some further discussion since it leads to the paradoxes already mentioned.

We are now in a position to recognize Haag's seminal contribution [Haag, 1962]: the *tacit* assumption of the irreducibility of the representation of the field algebra is *untenable.* Giving up this assumption allows one to resolve the paradoxes: Δ and hence the coefficients u and v in the Bogoliubov–Valatin transformation (77) — rather than being multiples of the identity — now belong to the *non-trivial* center $\mathcal{Z}$ of the representation canonically associated by the GNS construction corresponding to the equilibrium state of the system. The gauge group now acts in a non-trivial manner on $\mathcal{Z}$ and thus restores the symmetry of the theory. And in the limit considered, the time-evolution is well defined as an automorphism group of the von Neumann algebra generated by the representation. These technical niceties have been succesively refined — and confirmed — in subsequent investigations, cf. e.g. [Emch and Guenin, 1966; Thirring and Wehrl, 1967; Thirring, 1968; Dubin and Sewell, 1970; Sewell, 1982b].

5.2 Beyond Fock space: The Bose gas

Even before the modular structures were formally recognized by mathematicians, their first instantiation appeared in QSP. One can indeed discern these structures in the pioneering re-examination Araki and Woods [1963] made of the Bose–Einstein model for an ideal quantum gas; for the original version of the model, cf. subsection 2.4. The present subsection summarizes the principal aspects of the Araki–Woods treatment.

The reader is assumed to be familiar with the definition of the Weyl form of the canonical commutation relations (CCR) for a countably infinite number of degrees of freedom, as a family $\{W(f) \mid f \in \mathcal{D}(\mathbb{R}^3)\}$ of unitary operators acting on the (boson) Fock space $\mathcal{F} := \bigoplus_{N=0}^{\infty} {}^s\mathcal{H}^N$ and satisfying $\forall\, f,g \in \mathcal{D}(\mathbb{R}^3)$: $W(f)\,W(g) = \exp\{-i\,\mathrm{Im}(f,g)/2)\}$; where $\mathcal{D}(\mathbb{R}^3)$ is the space of all infinitely differentiable functions $f : \mathbb{R}^3 \to \mathbb{C}$ which have compact support; and ${}^s\mathcal{H}^N$ is the symmetric $N-$fold tensor product of the one-particle space $\mathcal{H}^1 = \mathcal{L}^2(\mathbb{R}^3)$ with itself; cf. e.g. [Emch, 1972a], or [Halvorson, 2006].

For the Bose gas at temperatures $T > T_c$ where T_c is the critical temperature found by Bose and Einstein, the GNS representation π_g corresponding to the gaseous normal phase — in the thermodynamical limit at fixed density ρ and chemical activity z — is given as follows. The Hilbert space of the representation π_g may be identified with $\mathcal{H} = \mathcal{F} \otimes \mathcal{F}$; its cyclic vector is $\Phi = \Phi_o \otimes \Phi_o$, where Φ_o is the vacuum vector in $\mathcal{F}$. Then

(80) $\pi_g[W(f)] = W(\zeta_+ f) \otimes W(K\zeta_- f)$

where completeness demands that we specify that $(\zeta_+ f)\tilde{}\,(k) = [1+\rho(\beta,z;k)]^{\frac{1}{2}}\,\tilde{f}(k)$, and $(\zeta_- f)\tilde{}\,(k) = [\rho(\beta,z;k)]^{\frac{1}{2}}\,\tilde{f}(k)$, $(Kf)\tilde{}\,(k) = \tilde{f}(k)^*$; $\rho(\beta,z;k) = z[\exp(\beta\epsilon(k)) - z]^{-1}$ with $\epsilon(k) = |k|^2/2m$ and z is determined by ρ and β through $\rho = (2\pi)^{-3}\int d^3k\ \rho(\beta,z;k)$.

The von Neumann algebra $\mathcal{N}_g = \{\pi_g[W(f)] \mid f \in \mathcal{D}(\mathbb{R}^3)\}''$ is a factor, the commutant of which is $\mathcal{N}_g{}' = \{\nu_g[W(f)] \mid f \in \mathcal{D}(\mathbb{R}^3)\}'$ where

(81) $\nu_g[W(f)] = W(K\zeta_- f) \otimes W(\zeta_+ f)$.

Note that ν_g also gives a representation of the Weyl CCR.

In what I believe was the first presentation of the programme proposed in [Haag *et al.*, 1967] to a wide audience of mainstream physicists, namely the huge IUPAP 1966 Copenhagen meeting on statistical mechanics, Winnink [Winnink, 1967] started indeed with a summary of the above results. As the duality between the von Neumann algebra and its commutant is already a property of finite systems — see scholium 33 and remark 34 above — Winnink's emphasis was that this property may persist in general for systems endowed with infinitely many degrees of freedom, as is the case in this specific model — the Bose gas — where the thermodynamical limit of canonical equilibrium is controlled. The emphasis on dealing with infinite systems — also advocated in the lecture [Verboven, 1967] preceding Winnink's — raised eyebrows with many of the physicists in the Copenhagen audience, to wit: "Wouldn't one think that, so to say, the motivation of going to an infinite system would be to obtain simpler results than are obtained for a finite system?" [Uhlenbeck, 1967]; or even more pointedly: "What does this have to do with statistical mechanics?"[van Kampen, 1967]. The conjecture was already floated that the formalism could be useful for an adequate description of phase transitions, a conjecture I will examine in subsections 5.6 and 5.7.

In retrospect, it is quite remarkable that Araki and Woods [1963] had already unearthed several features that were later placed in the context of the general theory that was to be built on the subsequent work of Tomita and Takesaki [Tomita, 1967; Takesaki, 1970a] for the mathematical formalism and the work of Haag *et al.* [1967] for its application to QSP. Among the results by Araki and Woods, one may note that the von Neumann factor $\mathcal{N}_g$ they constructed for $T > T_c$ is of type III — a type of factor the existence of which was known, but for which examples were then quite elusive even in the pure mathematics literature — and this was the first occurrence of this type of factor in QSP, although their ubiquity was later recognized all over in QSP and in QFT; and also in pure mathematics, but that is another story. In addition, Araki and Woods established that the unitary operators implementing time-evolution and space-translation on the von Neumann algebra $\mathcal{N}_g$ do not belong to this algebra. They also discussed the representations relative to the superfluid phase which occurs for temperatures $0 < T < T_c$, and they found that the associated GNS representation is an integral of factor representations. Incidentally, they do mention that this points to a formal analogy with the mathematical structure Haag found in his study of the BCS model; see subsection 5.1 above.

5.3 The KMS condition and the Heisenberg model

The first proof that the KMS conditions themselves are actually satisfied in concrete infinite quantum systems was provided by Araki [1969] for a class of one-dimensional quantum spin-lattice models which includes the archetypal model — originally proposed by Heisenberg [1928] as a putative model for ferromagnetism — defined by the local, so-called "exchange" Hamiltonian:

$$(82)\quad H_\Lambda = -J \sum_{k=a}^{b-1} \sigma_k \cdot \sigma_{k+1}$$

where J is the coupling constant describing interactions of neighbouring quantum spins $\sigma_k = ({\sigma_k}^x, {\sigma_k}^y, {\sigma_k}^z)$ sitting on a regular, one-dimensional finite string $\Lambda =$

$[a, b] \subset Z^1$; and $\sigma_k \cdot \sigma_{k+1}$ denotes ${\sigma_k}^x {\sigma_{k+1}}^x + {\sigma_k}^y {\sigma_{k+1}}^y + {\sigma_k}^z {\sigma_{k+1}}^z$.

The problem of determining whether this quantum model would support ferromagnetism in its thermodynamical limit — even in this one-dimensional version — turned out to be much harder to handle than the classical Ising model where only interactions $J {\sigma_k}^z {\sigma_{k+1}}^z$ are considered.

For the classical models, a method known as the *transfer-matrix* — and actually proposed for the two-dimensional Ising model [Kramers and Wannier, 1941] — allows one to treat the one–dimensional version of this classical model in a few lines for nearest neighbour interactions, or even with strictly finite-range interactions, i.e. when the interactions are strictly zero between spins that are further apart than a finite distance (the same for all pairs). At the cost of quite some work [Ruelle, 1968b], the method can be made to work for interactions that extend to infinity, while decaying sufficently fast so as to have finite moment or so that the surface energy has a bound independent of the volume.

As emphasized in some detail in [Emch, 1972b], even the nearest neighbour quantum Heisenberg model requires an extension akin to the method used for the infinite range classical case. Araki [1969] thus managed to control the thermodynamical limit φ of the canonical equilibrium state and its time correlation functions well enough to establish that for all positive temperatures $\beta > 0$, the state φ satisfies the KMS condition; and that it is extremal with respect to this condition — i.e. cannot be decomposed into a mixture of states satisfying the KMS condition — and shows no spontaneous magnetization. Thus, while physicists — with their understanding of the onset of collective behaviour in the classical case — could anticipate that this quantum one-dimensional model would not exhibit any ferromagnetic phase transition, Araki proves it.

The class of models for which Araki established the above results is strongly dependent on the one-dimensionality of the "lattice" $\mathbb{Z}$. Nevertheless, the proof does not require that the interactions be isotropic, i.e. couplings between the different components of the spins do not need to be the same in all directions. Moreover the proof does not require that the interactions between the spins on the lattice be restricted to nearest neighbours: in the original version of the proof, it was only required that they vanish between spins that are further apart than a fixed (but arbitrary) finite distance, but even this restriction has been relaxed to cover the same range as the corresponding classical models. Finally, whereas in the Heisenberg model the individual half-spins are described by copies of the algebra $\mathcal{M}(2, \mathbb{C})$, the proof accommodates as well the cases where $\mathcal{A}_k \simeq \mathcal{M}(n, \mathbb{C})$ with $n < \infty$.

Hence Araki's results support a conjecture that pure thermodynamical phases may be described by extremal KMS states; see subsection 5.7 below for further evidences.

It may be added that if, in addition, lattice translation invariance is built into the theory through the local Hamiltonian H_Λ, e.g. as in (82), then the limiting KMS state φ is invariant under the group $\mathbb{Z}$ of the lattice translations, and — since its GNS representation leads to a factor — φ is also extremal with respect to this condition, so that space-correlations between spins decay very fast as their distance increases. Here, technically speaking, φ is exponentially, uniformly clustering in space; i.e. for any quasi-local observable $A \in \mathcal{A}_o$, there exist positive constants γ and δ such that for all finite N and all $B \in \mathcal{A}_{\mathbb{Z}\setminus[-N,N]}$: $|\varphi(AB) - \varphi(A)\varphi(B)| \leq \delta \, \|B\| \, \exp(-\gamma N)$.

Before leaving the Heisenberg model, note that in the case $T = 0$, it also offers a very nice toy model for QFT; cf. e.g. [Streater, 1967].

5.4 The KMS condition and stability

The following five points summarize the KMS story I have told so far.

1. Von Neumann's definition of quantum canonical states at finite temperature is limited to finite systems; and this limitation renders cumbersome (at best) the formalism's application to QSP — see subsection 3.3.

2. For finite systems, the von Neumann equilibrium states are exactly those that satisfy a formal analytic condition, the so-called KMS condition — see subsection 4.1.

3. The KMS condition can be extended beyond the mathematical formalism laid down by von Neumann — see subsection 4.2.

4. The KMS condition is satisfied in some concrete models of infinite systems for states that have a reliable interpretation as temperature equilibrium states — see subsections 5.2 and 5.3.

5. The KMS condition appears as well in a purely mathematical context, the Tomita–Takesaki theory of modular algebras which turns out to be very fertile; while the latter aspect of the story would carry us beyond the bounds of this essay, some hints are briefly alluded to in see subsections 4.2 and 5.2.

Before the KMS theory could be deemed adequate as a physical theory, it ought to meet at least two more concerns: (i) the formalism should allow the mathematical description of quantum phenomena that escape the grip of von Neumann's formalism; (ii) the KMS states should be stable. Subsections 5.2 and 5.3 above indicate how the first of these two concerns is met; further examples will be presented in subsection 5.7. The present subsection addresses the second concern, as various stability criteria — labeled A to E — are discussed. The order of the presentation is to direct the reader's attention to the progressive emergence of formulations whereby KMS states are characterized in terms increasingly germane to those of variational principles.

A. Cut-and-paste stability.

We begin with a model that is sufficiently simple to provide exact results supporting the expectation that a large system in a canonical equilibrium state should be able to serve as a *thermal reservoir* for "any" of its parts. The model is a variation on the theme of the so-called XY-model; this variation was proposed and solved in [Emch and Radin, 1971]; further references will be given at the end of this subsection.

The X-Y model itself — referred to below as the 'un-partitioned system' — is a one-dimensional quantum spin-lattice gas with finite-range interactions. Specifically, for any region $\Lambda = \{k \in \mathbb{Z} \mid a \leq k \leq b\}$ with $-\infty < a+1 < b < \infty$, the Hamiltonian is

$$(83)\quad H_{[a,b]} = -\sum_{k=a}^{b-1} (1+\zeta)\sigma_k{}^x\sigma_{k+1}{}^x \;+\; (1-\zeta)\sigma_k{}^y\sigma_{k+1}{}^y \quad .$$

From the work of Araki — see subsection 5.3 above — we learned that the thermodynamical limit (as $a \to -\infty$, $b \to \infty$) of both of the following objects exists: (i) the canonical equilibrium state $\varphi_{[a,b]}$ at any finite natural temperature $\beta > 0$ and (ii) the time-evolution $\alpha_{[a,b]}$; and that the resulting state φ and evolution α of the infinite system satisfy the KMS condition.

We now split the total system in two non-interacting parts: a finite region to which we affix the subscript S, and an infinite region to which we affix the subscript R, which is the complement of Λ_S in $\mathbb{Z}$, namely:

$$\Lambda_S = [c,d] \quad \text{and} \quad \Lambda_R = (-\infty, c-1] \cup [d+1, \infty)$$

with $-\infty < a < c-1\,;\; c < d-1\,;\; d+1 < b < \infty$.

This partitioned system can be viewed as the thermodynamical limit of a finite system with Hamiltonian:

$$(84)\quad \tilde{H}_{[a,b]} = H_{[a,c-1]} + H_{[c,d]} + H_{[d+1,b]} \quad .$$

Clearly the C^*-algebras for both the original system and the partitioned system are the same, namely the C^*-inductive limit $\mathcal{A} := \otimes_{k\in\mathbb{Z}} \mathcal{A}_k$ where the $\mathcal{A}_k$ are copies of the algebra $\mathcal{M}(2,\mathbb{C})$ of the 2×2 matrices with complex entries. Thus $\mathcal{A} = \mathcal{A}_S \otimes \mathcal{A}_R$ where $\mathcal{A}_S := \otimes_{k\in\Lambda_S} \mathcal{A}_k$ and $\mathcal{A}_R := \otimes_{k\in\Lambda_R} \mathcal{A}_k$.

Again, as for the original (un-partitioned) system, the thermodynamical limit of the canonical equilibrium state and of the evolution of the partitioned system, defined from (84), do exist and satisfy the KMS condition; they are denoted here by $\tilde{\varphi}$ and $\tilde{\alpha}$.

Note that φ and $\tilde{\varphi}$ are different. For instance, φ is invariant with respect to translations along the chain, while $\tilde{\varphi}$ is not. Nevertheless, a first stability property of this model is established in [Emch and Radin, 1971], namely:

$$(85)\quad \forall\, A \in \mathcal{A} \,:\; \lim_{|t|\to\infty} \tilde{\varphi}(\alpha_t[A]) = \varphi(A) \quad .$$

Hence, as the full evolution α unfolds, the correlations between S and R that were cut by the partitioning are re-established: the partition is erased.

Moreover, let $\tilde{\varphi}_S$ denote the restriction of $\tilde{\varphi}$ to $\mathcal{A}_S$; and similarly for R; one has $\tilde{\varphi} = \tilde{\varphi}_S \otimes \tilde{\varphi}_R$. The evolution $\tilde{\alpha}$ preserves the partitioning, namely $\forall A \in$

$\mathcal{A}_S$ [resp.$\mathcal{A}_R$] : $\tilde{\alpha}_t[A] \in \mathcal{A}_S$ [resp.$\mathcal{A}_R$]. Hence, we have $\tilde{\alpha} = \tilde{\alpha}_S \otimes \tilde{\alpha}_R$; i.e. the two systems evolve independently. Again the KMS conditions are satisfied for S and R separately.

After this partitioning, let us now change the temperatures of S and R so that (i) $\tilde{\varphi}_{S,\beta_S}$ is the canonical equilibrium on $\mathcal{A}_S$ at some temperature β_S with respect to the evolution $\tilde{\alpha}_S$; and similarly (ii) with β_R (possibly different from β_S) for $\tilde{\varphi}_{R,\beta_R}$ (w.r.t $\tilde{\alpha}_R$) on $\mathcal{A}_R$. Let thus $\tilde{\varphi}_{S,\beta_S} \otimes \tilde{\varphi}_{R,\beta_R}$ be the initial state of the partitioned system; and denote by φ_β the canonical equilibrium on the whole system, at temperature β with respect to the original interacting evolution α. Then, the following is proven [Emch and Radin, 1971] for all β_S, $\beta_R > 0$ and for all $A \in \mathcal{A}$:

$$(86) \quad \lim_{|t|\to\infty} \tilde{\varphi}_{S,\beta_S} \otimes \tilde{\varphi}_{R,\beta_R}(\alpha_t[A]) = \varphi_\beta(A) \ \text{ with } \ \beta = \beta_R \ .$$

Hence the name *'cut-and-paste stability'.* The system is first partitioned in two parts, a finite system S surrounded by an infinite system R that do not interact with one another: the interactions across the boundaries [i.e. between the sites $c-1$ and c; and between the sites d and $d+1$] have been 'cut'. In this configuration, the finite system S and the infinite system R are separately put at (different) temperatures β_S and β_R. When these systems are 'pasted' back together, one finds that the joint evolution drives the full system $S \cup R$ to a temperature β which has to be the temperature at which R was initially, namely $\beta = \beta_R$. In this sense, R serves as a thermal reservoir for S.

The special property of the model responsible for this result is that it satisfies a remarkable condition which I now describe.

Let γ be the automorphism of $\mathcal{A}$ uniquely determined by

$$(87) \quad \forall\, k \in Z \ : \ \begin{cases} \gamma[{\sigma_k}^z] = & {\sigma_k}^z \\ \gamma[{\sigma_k}^x] = - & {\sigma_k}^x \\ \gamma[{\sigma_k}^y] = - & {\sigma_k}^y \end{cases} \quad .$$

Note in particular that the original Hamiltonian $H_{[a,b]}$ and the cut Hamitonian $\tilde{H}_{[a,b]}$ belong to the even algebra $\mathcal{A}_e := \{A \in \mathcal{A} \mid \gamma[A] = A\}$. This entails that in the thermodynamical limit $\varphi \circ \gamma = \varphi$ and $\gamma \circ \alpha \circ \gamma = \alpha$; and similarly for all the corresponding objects obtained after the partitioning. In particular, the evolution preserves the even algebra, i.e. $\forall\, (t, A) \in \mathbb{R} \times \mathcal{A}_e : \alpha_t[A] \in \mathcal{A}_e$.

Now, the special property of the model can be explicitly stated:

$$(88) \quad \forall\, A, B \in \mathcal{A}_e \ : \ \lim_{|t|\to\infty} \|A\alpha_t[B] - \alpha_t[B]A\| = 0 \quad ;$$

i.e. the evolution, when restricted to the even observables of the model, is *strongly asymptotically abelian.*

The proof — and an immediate generalization — of (86) above is a direct consequence of quantum ergodic theory (see subsection 3.5). First, one notices that φ is uniformly clustering in space, i.e. for every $\epsilon > 0$ and every $A \in \mathcal{A}$ there exists a finite region Λ such that for every B outside this region $|\varphi(AB) - \varphi(A)\,\varphi(B)| \leq \epsilon\|B\|$. This entails that the KMS state φ is extremal with respect to this condition, i.e. cannot be decomposed into a convex combination of other KMS states; see subsection 5.6 below, and in particular definition 57. These properties are inherited by the restriction φ_e of φ to the even algbra $\mathcal{A}_e$. The asymptotic abelianness of the evolution then implies [Araki and Miyata, 1968] that the state φ_e is not only time-invariant — as we know every KMS state must be — but it is also extremal with respect to this condition, i.e. cannot be decomposed into other time-invariant states, which is to say that φ_e cannot be written as $\varphi_e = \lambda\psi_e + (1-\lambda)\chi_e$ with $0 < \lambda < 1$ and ψ_e, χ_e time-invariant, *unless* $\psi_e = \chi_e = \varphi_e$.

Since $\varphi, \varphi_e, \tilde{\varphi}_S, \tilde{\varphi}_R$ are even, one looses no information by carrying out the proof of (85) and (86) in $\mathcal{A}_e$; in particular, (88) implies the existence of the point-wise limit of $(\varphi_S \otimes \varphi_R) \circ \alpha_t$ in the LHS of (86); then the above argument shows that it must coincide with φ.

Note further that what is proven in [Emch and Radin, 1971] is in fact a stronger result, which implies (86) and thus (85) as particular cases, namely that for all even states ψ_S of the system S :

$$(89)\quad \forall\, A \in \mathcal{A} \; : \; \lim_{|t|\to\infty} \psi_S \otimes \tilde{\varphi}_R(\alpha_t[A]) = \varphi_\beta(A)$$

which therefore reinforces the 'cut-and-paste stability' interpretation proposed immediately after equation (86).

This result may be further generalized in two ways. First, the restriction that ψ_S in (89) be an even state can be dispensed with; cf. [Araki and Barouch, 1982]. Second, as was already noticed in [Emch and Radin, 1971], an ergodic or averaged version of (89), specifically, with η denoting an invariant mean on the group $\mathbb{R}$:

$$(90)\quad \forall\, A \in \mathcal{A} \; : \; \eta\{\psi_S \otimes \tilde{\varphi}_R(\alpha[A])\} = \varphi_\beta(A)$$

obtains [Emch and Radin, 1971], even if only a weaker version of (88) holds, namely the condition of $\eta-$asymptotic abelianness (60), i.e.

$$(91)\quad \forall\, A, B, C \in \mathcal{A}_e \; : \; \eta\{\varphi(C^*\,[A\,\alpha[B] - \alpha[B]\,A]\,C)\} = 0 \quad .$$

Depending on one's intellectual temperament, *either* the general argument presented earlier, *or* the specific model just reviewed, raises the question of whether the KMS condition could be derived from some general stability argument. This question is addressed from several angles in (B)–(E) below.

The model presented above was discussed again in [Robinson, 1973]; see also [Araki and Barouch, 1982] and references therein. It belongs to a long line of inquiries that started as attempts to derive Newton's cooling law from first principles; related problems are usually understood under the generic label 'return to equilibrium'. As of this writing, the latest comprehensive entry on the subject may be [Bach *et al.*, 2000] where a wealth of "novel technical devices" are brought to bear; the reader will also find there an informative sample of the large literature on the subject. In a broad sense, several — but not all — of the criteria of stability in this subsection also address this perennial problem of return to equilibrium from small or local deviations. Its ubiquity however should not overshadow two other important and largely unsolved problems; cf. subsection .6.4 below.

B. Stability against local perturbations.

Various conditions of asymptotic abelianness were investigated by Kastler *et al.* For a summary, see [Kastler, 1976] which also offers a annotated bibliography. For their main stability theorem, they settled on the notions described in definitons 41 and 42 below.

DEFINITION 41. Let $\mathcal{A}$ be a C$^*-$algebra. An evolution $\alpha : \mathbb{R} \to \mathrm{Aut}(\mathcal{A})$ is said to be L^1- asymptotically abelian on a norm dense $^*-$subalgebra $\mathcal{A}_o \subset \mathcal{A}$ when

$$\forall\,(t, A) \in \mathbb{R} \times \mathcal{A}_o : \alpha_t[A] \in \mathcal{A}_o\,;$$

and

$$\forall\, A, B \in \mathcal{A}_o \;:\; \int_{-\infty}^{+\infty} dt\, \|B\,\alpha_t[A] - \alpha_t[A]\,B\| < \infty \quad .$$

Some preliminary notations are required for Definition 42 below. With $\mathcal{A}$ and α as in definition 41 let $\mathcal{A}_{sa} = \{A \in \mathcal{A} \mid A = A^*\}$, and let $\mathcal{S}$ be the set of all states on $\mathcal{A}$, equipped with its weak topology. For $\varphi \in \mathcal{S}$ and an element $h \in \mathcal{A}_{sa}$ with $\varphi(h^2) > 0$, define

(i) the perturbed state φ^h by $\varphi^h : A \in \mathcal{A} \to \frac{1}{\varphi(h^2)}\,\varphi(hAh) \in \mathcal{A}$;

(ii) the perturbed evolution $\{{\alpha^h}_t \mid t \in \mathbb{R}\}$ by ${\alpha^h}_t : A \in \mathcal{A} \mapsto {U^h}_t\,\alpha_t[A]\,{{U^h}_t}^*$ where $\{{U^h}_t \mid t \in \mathbb{R}\}$ satisfies the so-called 'co-cycle differential equation' (the derivative is w.r.t. the norm-topology)

$$\forall\, t \in \mathbb{R} \;:\; i\,\tfrac{d}{dt}{U^h}_t = {U^h}_t\alpha_t[h] \;\text{ with initial condition } {U^h}_o = I\,.$$

To understand the sense in which α^h may be viewed as the perturbed evolution corresponding to h, note that the above co-cycle equation admits a unique continuous solution $t \in \mathbb{R} \mapsto {U^h}_t \in \mathcal{A}$; it can be computed explicitly as the norm-convergent Dyson series:

$${U^h}_t = \sum_{n=0}^{\infty} {C^h}_{t,n} \text{ with } {C^h}_{t,n} = (-i)^n \int_o^t dt_n \int_o^{t_n} dt_{n-1} \ldots \int_o^{t_2} dt_1 \alpha_{t_1}[h] \ldots \alpha_{t_n}[h].$$

This solution satisfies: (i) ${U^h}_t$ unitary, and

$$(ii) \qquad \forall\, s, t \in \mathbb{R} \;:\; {U^h}_{s+t} = {U^h}_s\,\alpha_s[{U^h}_t]\,.$$

Consequently, the evolution defined as $\{{\alpha^h}_t \mid t \in \mathbb{R}\}$ is a group of automorphisms of $\mathcal{A}$ with, in particular:

$$\forall\, s, t \in \mathbb{R} \;:\; {\alpha^h}_{s+t} = {\alpha^h}_s \circ {\alpha^h}_t \quad .$$

The interpretation of α^h as the evolution resulting from the perturbation of α by the operator h obtains from the following relation between the generators of α^h and α :

$$i\frac{d}{dt}{\alpha^h}_t\,|_{t=0} = i\frac{d}{dt}\alpha_t\,|_{t=0} + \delta^h \quad \text{with} \quad \delta^h : A \in \mathcal{A} \mapsto [h, A] := hA - Ah \in \mathcal{A} \quad .$$

DEFINITION 42. With the above notations, an $\alpha-$invariant state φ on $\mathcal{A}$ is said to be stable against inner perturbations, whenever there is a neighbourhood $\mathcal{V}_\varphi \subset \mathcal{S}$ of φ such that $\forall\, A \in \mathcal{A}$, and $\forall\, h \in \mathcal{A}_{sa}$ with $\varphi^h \in \mathcal{V}_\varphi$:

1. $\forall\, t \in \mathrm{I\!R} :\ \varphi^h({\alpha^h}_t[A]) = \varphi^h(A)$;
2. with $\lambda \in \mathrm{I\!R} :\ \lim_{\lambda\to 0} \varphi^{\lambda h}(A) = \varphi(A)$;
3. $\lim_{t\to\infty} \varphi^h(\alpha_t[A]) = \varphi(A)$.

THEOREM 43. *With* $\mathcal{A}$, $\mathcal{A}_o$ and α *as in definition 41, assume that* α *is* L^1-*asymptotically abelian on* $\mathcal{A}_o$. *Let* φ *be an* $\alpha-$*invariant state on* $\mathcal{A}$ *and assume that* φ *is stable against inner perturbations in the sense of definition 42. Then — under three ancillary conditions to be discussed below —* φ *satisfies the KMS condition with respect to* α *for some natural temperature* β .

REMARKS 44. The ancillary conditions of the theorem are sketched in the three entries below.

1. The state φ is assumed *not* to be a trace, i.e. there exist $A, B \in \mathcal{A}$ such that $\varphi(AB) \neq \varphi(BA)$. This is meant to avoid the classical circumstance that would arise in the limit of infinite temperature, i.e. $\beta = 0$, i.e. $T = \infty$.

2. In the GNS representation canonically associated to φ, the generator of the unitary group $U(\mathrm{I\!R})$ that implements $\alpha(\mathrm{I\!R})$ is assumed *not* to be one-sided. This is meant to avoid the opposite circumstance where φ would be a zero-temperature ground state, i.e. $\beta = \infty$, i.e. $T = 0$.

3. The state φ is assumed to be hyperclustering of order 4 on the $^*-$subalgebra $\mathcal{A}_o$. This technical condition requires the following to hold: for every positive integer $p \leq 4$ and all $A_1, \dots, A_p \in \mathcal{A}_o$, there exist positive constants C and δ such that

   (92) $\forall\, t_1, \dots, t_p \in R :\ \varphi_p^T(\alpha_{t_1}[A_1] \cdots \alpha_{t_p}[A_p]) \leq C\,\{\,1 + \max\,|t_k - t_l|^{1+\delta}\,\}^{-1}$

   where the truncated correlations φ_p^T are defined recursively by $0 = \varphi_o^T$, $\varphi(A) = \varphi_1^T(A)$ and $\varphi(A_1, \dots, A_p) = \sum_P \varphi_{n_1}^T(A_{k_1}, \dots, A_{k_{n_1}}) \dots \varphi_{n_j}^T(A_{q_1}, \dots, A_{q_{n_j}})$ and the sum carries over all order-preserving partitions of $S = \{1, 2, \dots, p\}$ in subsets $S_j \subseteq S$ satisfying the following conditions: $S = \cup_j S_j$, $j \neq k \Rightarrow S_j \cap S_k = \emptyset$, and within each $S_j = \{k_1, k_2, \dots, k_{n_j}\}$: $k_1 < k_2 < \dots < k_{n_j}$. The reader will verify immediately that $\varphi(A_1, A_2) =$

$\varphi_2^T(A_1, A_2) + \varphi_1^T(A_1)\,\varphi_1^T(A_2)$, and then realize that the recursion relation explains better what is going on with higher truncated correlations than writing explicitly the summations over P.

Note that the φ_p^T provide a hierarchy where all correlations of lower order already have been taken into account. In particular in the case of the CCR, a remarkable result of Robinson [Robinson, 1965] shows that either this hierachy goes up indefinitely or, if the truncated φ_n^T vanish for all $n \geq N$ with $N > 2$, then they must vanish for all $n > 2$.

The concept of truncated φ_n^T is not a stranger. It comes to us as a quantum cousin of the "cumulants" of classical probability theory and of the "Ursell functions" of classical statistical mechanics. The classical equivalent of Robinson's theorem gives a characterization of the Gaussian distribution, which translates in quantum statistics as yet another characterization of the canonical equilibrium state of an assembly of free harmonic oscillators. Robinson's theorem thus gives a foretaste of why it is so difficult to produce and/or control models of QFT and QSP that are not "quasi-free".

To sum up, the third ancillary condition of the theorem aims to convey that in the course of time all time-correlations of order $p \leq 4$ are to decay rapidly enough for long time separations.

The investigations by Kastler *et al.* reported above appear to be systematically predicated on conditions of time-asymptotic abelianness (definition 41) and time-hyperclustering (remark 44(3)). Thus compare these with any of the conditions encountered in sections 3 and 4; the latter are naturally satisfied for space translations, but in constructing specific models, even these conditions are extremely difficult to impose straight on the microscopic dynamics, i.e. on the Hamiltonian that is to describe the time evolution. Whether this is an intrinsic shortcoming of the theory behind theorem 43 above, or an indication of some lack of either imagination or technical dexterity on the part of model builders remains open at this stage. Nevertheless, it appears that one weak form of asymptotic abelianness is not only sufficient but also necessary when one wants to identify, among KMS states, those that are merely extremal with respect to this condition, from those that are, moreover, extremal with respect to time-invariance; cf. e.g. [Emch, 1972a, corollary 2, p. 206]; or remark 63(6) below. Here again, reminiscences from the perennial ergodic dreams in classical statistical *Hamiltonian mechanics* would incline some to hope that such an identification could perhaps be in the cards. As I have recognized in several other parts of this essay, my crystal ball remains clouded on this issue.

C. Thermal reservoir stability.

Consider the intuitive idea that a system R may be construed as a "thermal reservoir" at temperature β, if it drives suitably devised test systems S to equilibrium at temperature β when they are coupled to R. Kossakowski *et al.* [1977] proposed to formalize this idea in the following manner; see also [Sewell, 2002, pp. 114–116].

For a concrete motivation, compare the specific XY-model described in part **A** of this subsection.

To model situations where one expects that R ought to be very much larger than S in order to exclude feedbacks from the test system S onto the reservoir R, one assumes that R is infinite and S is finite.

The putative reservoir R is described by a triple $\{\mathcal{A}^R, \alpha^R, \varphi^R\}$ where $\mathcal{A}^R$ is a C^*-algebra; α^R is an evolution group of automorphisms of $\mathcal{A}^R$; and φ^R is a state on $\mathcal{A}^R$, invariant under the evolution α^R. Denote by $\delta^R := i\frac{d}{dt}{\alpha^R}_t\big|_{t=o}$ the generator of the evolution α^R. Some ancillary conditions on R will be specified later.

The test system S is a dynamical system in the sense of von Neumann, i.e. is described by: $\{\mathcal{A}^S, \alpha^S, \varphi^S_\beta\}$ where $\mathcal{A}^S = \mathcal{B}(\mathcal{H})$; α^S is the evolution generated by a Hamiltonian H^S such that for all temperatures $\beta > 0$, $Z := \text{Tr}\exp(-\beta H^S) < \infty$; and φ^S is given by

$$(93)\quad \varphi^S_\beta(A^S) = \text{Tr}\,\rho^S_\beta\, A^S \quad \text{with} \quad \rho^S_\beta = Z^{-1} e^{-\beta H^S} \quad .$$

$\delta^S = i\,[H^S\,,\,\cdot\,]$ will denote the generator of α^S. Finally, $\mathfrak{S}^S$ will denote the set of all countably additive states on $\mathcal{A}^S$.

A family $\{\alpha^\lambda \mid \lambda \geq 0\}$ of dynamical couplings between R and S is described by groups of automorphisms on $\mathcal{A} = \mathcal{A}^R \otimes \mathcal{A}^S$, the generator of which is of the form:

$$(94)\quad \left.\begin{array}{c} \delta^\lambda = \delta^R \otimes I + I \otimes \delta^S + \lambda\delta_V \quad \text{where} \\ \delta_V : A \in \mathcal{A} \,\mapsto\, i\,\lambda[V\,,\,A] \in \mathcal{A}, \quad \text{with} \quad V \in \mathcal{A}_{sa} \end{array}\right\} \quad .$$

As the ancillary conditions on R are specified, so will be the form of V; see (97) and (98) below.

The next step in the modeling is devised to emphasize the sense in which the long-time cumulative effects on S of the evolution α^λ are accounted for when R and S are coupled. For this Kossakowski *et al.* [1977] appeal to the so-called van Hove limit, an instance of which already appeared in subsection 3.5; see also remark 45 below. For the system at hand here, the van Hove limit takes the following form. First, it considers only a reduced evolution, namely only what the system S experiences of the total evolution; mathematically this reduction is achieved by $E : \mathcal{A} \to \mathcal{A}_S$, the conditional expectation defined, for all $A_S \otimes A_R$, by $E[A_S \otimes A_R] = A_S\, \varphi_R(A_R)$, and then extended by linearity and continuity to $\mathcal{A}$. Secondly, the van Hove limit requires to focus on a long-time/weak-coupling regime defined by rescaling time with an inverse power of the interaction strength. Thus, the van Hove limiting procedure consists here in proving that the following limit exists for all positive 'rescaled' times s :

$$(95)\quad \gamma^S_s : A_S \in \mathcal{A}^S \,\mapsto \lim_{\substack{\lambda\to 0\;;\;t\to\infty \\ s=\lambda^2\,t}} \alpha^S_{-t} \circ E \circ \alpha^\lambda_t[A_S] \,\in\, \mathcal{A}^S \quad .$$

REMARKS 45. This type of limit has a long history. I learned it first from van Hove [van Hove, 1955] where the author had proposed it as a tool to relate macroscopic transport phenomena to the microscopic dynamics that is expected to underlie them. It emphasizes that in such discussions time ought to be rescaled in a way determined by the strength λ of the interaction. Some justifications for taking such a limit will be discussed in subsection 6.1 below.

Finally, given two C^*- algebras $\mathcal{A}$ and $\mathcal{B}$, and n a non-negative integer, one says that a map $\gamma : \mathcal{A} \to \mathcal{B}$ is *n-positive* whenever it is linear, and the induced map $\gamma_n : \mathcal{A} \otimes \mathcal{M}(n,\mathbb{C}) \to \mathcal{B} \otimes \mathcal{M}(n,\mathbb{C})$ is positive, i.e. the image of any positive element in $\mathcal{A} \otimes \mathcal{M}(n,\mathbb{C})$ is a positive element in $\mathcal{B} \otimes \mathcal{M}(n,\mathbb{C})$. When either $\mathcal{A}$ or $\mathcal{B}$ is abelian, a positive map is necessarily $n-$positive; hence $n-$positivity is a notion new to the non-commutative context of QSP. Furthermore, a map is said to be *completely positive* whenever it is $n-$positive for all $n \in \mathbb{Z}^+$. In connection with expressions like the right-hand side of (95) above, note that the composition of completely positive maps is again completely positive; and that automorphisms, states, injections and conditional expectations are completely positive maps. A collection $\{\gamma_s \mid s \in \mathbb{R}^+\}$ of maps of $\mathcal{A}$ into itself is said to form a semi-group whenever γ_o is the identity map, and $\forall\,(s,t) \in \mathbb{R}^+ \times \mathbb{R}^+ : \gamma_{s+t} = \gamma_s \circ \gamma_t$.

This should exhaust the list of general preliminaries necessary to describe the stability criterion proposed by Kossakowski *et al.* [1977], namely:

DEFINITION 46. A system $\{\mathcal{A}^R, \alpha^R, \varphi^R\}$ is said to be a thermal reservoir at temperature β whenever there is a "large enough" collection $\mathfrak{T}_\beta$ of test systems $\{\mathcal{A}^S, \alpha^S, \varphi^S_\beta\}$ and dynamical couplings $\{\alpha^\lambda\}$ such that

1. the van Hove limit (95) exists, and defines a semi-group of completely positive transformations $\{\gamma^S_s \mid s \in \mathbb{R}^+\}$ of $\mathcal{A}_S$;

2. the canonical von Neumann equilibrium state φ^S_β on $\mathcal{A}_S$ is the only state $\varphi \in \mathfrak{S}^S$ that is invariant under both α^S and γ^S;

(96) *3.* $$\forall\,(\psi^S, A^S) \in \mathfrak{S}^S \otimes \mathcal{A}^S : \lim_{s\to\infty} \psi^S\left(\gamma^S_s[A_S]\right) = \varphi^S_\beta(A_S) \quad .$$

The term "large enough" in the above definition admittedly needs to be made more precise: this is where the ancillary conditions on the interaction V and the reservoir R enter the picture and allow one to prove scholium 47 and theorem 48 below.

One condition is that the interaction V in (94) be of the form

$$(97) \quad V = \sum_{k=1}^{n} B^R_k \otimes B^S_k \quad \text{with} \quad \left\{ \begin{array}{l} n \text{ is finite} \\ B^R_k \in \mathcal{A}^R{}_{sa} \text{ and } \varphi^R(B^R_k) = 0 \\ B^S_k \in \mathcal{A}^S{}_{sa} \end{array} \right\} \quad .$$

Note that the conditional expectation $E[V]$ of V vanishes.

An additional condition is that there exists $\mathcal{A}^R_o \subseteq \mathcal{A}^R$ such that: (i) $\text{Span}\{\mathcal{A}^R_o \cup I\}$ (where I is the identity in $\mathcal{A}^R$) is norm dense in $\mathcal{A}^R$; (ii) for all $B^R_k \in \mathcal{A}^R_o$, the functions $t \mapsto \varphi^R(B^R_j \alpha^R_t[B^R_k])$ are in L^1; and (iii) the multi-time truncated correlations, for the state φ^R to be tested, satisfy

(98) $t_1 < \ldots < t_l$ with $|t_j - t_k| \to \infty \Rightarrow \{\varphi^R\}^T\,(\alpha^R_{t_1}[B^R_{t_1}] \cdots \alpha^R_{t_n}[B^R_{t_n}]) \;\to 0.$

Upon taking advantage of [Davies, 1974, theorem 2.3], the following results were obtained in [Kossakowski *et al.*, 1977]:

SCHOLIUM 47. These ancillary conditions are sufficient to imply that condition (1) in definition 46 is satisfied for all finite S.

This ensures that the collection $\mathfrak{T}_\beta$ of test systems will indeed be large enough.

THEOREM 48. *When the circumstances just outlined are realized, the following conditions are equivalent:*

1. *for some temperature β, the state φ^R is a KMS state on $\mathcal{A}_R$ with respect to the evolution α_R;*
2. *the system R, in the state φ^R, is a thermal reservoir for temperature β in the sense of definition 46 with "large enough" sharpened by scholium 47.*

REMARKS 49.

1. Hence, every test system S in $\mathfrak{T}_\beta$ is driven to equilibrium at temperature β by the reservoir R exactly when φ^R satisfies the KMS condition for this temperature.
2. As mentioned before, this result is largely model-independent, and does not involve, *at least explicitly,* any assumption of time asymptotic abelianness. Moreover, instead of a single model for which one can prove that a special infinite system in equilibrium serves as thermal reservoir for each of its finite parts, the present theorem characterizes a collection $\mathfrak{T}_\beta$ of test systems S for which the infinite system R serves as a thermal reservoir. Thus, the theorem is an improvement on the particular motivating model discussed in paragraph **A** above.
3. Yet, as [Kossakowski *et al.*, 1977] noticed, the decay of multi-time correlations (98) that enables their proposal to work is reminiscent of the similar conditions imposed by Kastler *et al.* in their result on stability against local perturbations; see (92) in paragraph **B** above.
4. From an empirical point of view, the theorem may be regarded as specifying a procedure to lift the notion of temperature in equilibrium QSP from finite systems to infinite systems.
5. Nevertheless, it must be noted that the circumstances under which γ^S is known to satisfy condition (1) of definition 46 and the decay of correlations in (98) do appear to involve some clustering properties that may limit the domain of applicability of the theorem to cases where $\pi_\varphi(\mathcal{A}_R)''$ is a factor, and where φ^R is extremal under both the KMS condition and the condition

of time-invariance. Hence, asymptotic abelianness enters less conspiciously here.

D. Passivity.

In [Pusz and Woronowicz, 1978] the authors noticed a property of KMS states which they called *passivity;* and they found ways to show that this property in turn entails the KMS property under assumptions that *do not* involve asymptotic abelianness in *time.*

Specifically, let $\{\mathcal{A}, \varphi, \alpha\}$ be a dynamical system where $\mathcal{A}$ is a C^*-algebra, φ be a state on $\mathcal{A}$, and $\{\alpha_t \mid t \in R\}$ is a one-parameter group of automorphisms of $\mathcal{A}$. Let then $D(\delta)$ denote the domain of the generator δ of the evolution α, i.e. $D(\delta)$ is the linear subspace of all $A \in \mathcal{A}$ such that the derivative $\delta[A] := i\frac{d}{dt}\alpha_t[A]$ exists.

Consider now the situation obtained by letting this system interact during a finite time-interval with an outside system, so that the effect of their interaction on the system of interest may be assumed to be described as the perturbed dynamics α^h satisfying the differential equations:

$$(99)\quad \forall\, A \in D(\delta)\ :\ \begin{cases} i\frac{d}{dt}{\alpha^h}_t[A] = {\alpha^h}_t[\delta[A] + [h_t, A]] \\ \\ {\alpha^h}_{t=0}[A] = A \end{cases}$$

where h is an element of $C^1_+(\mathbb{R}, \mathcal{A}_{sa})$, the set of all continuously differentiable functions, with compact support in $\mathbb{R}^+$ and taking their values in the self-adjoint part of $\mathcal{A}$. The system is thus an open system for all times t in the support of h, i.e. for all times when the perturbation h is actually in effect. The condition that the support of h be compact and contained in $\mathbb{R}^+$ ensures that, for all times $T > \sup\{t \in \mathbb{R} \,|\, h_t \neq 0\}$, the external conditons are as they were at time $t = 0$. The smoothness condition $h \in C^1$ on the time-dependence of the external pertubation is a mathematical convenience that is physically reasonable. Then

$$(100)\quad L^h_T(\varphi) := \int_o^T dt\ \ \varphi({\alpha^h}_t\,[\frac{d}{dt}h_t])$$

describes the energy transmitted to the system in the time interval [0, T] during which the system was under the influence of the external perturbation h.

DEFINITION 50. The state φ is said to be passive if for all $h \in C^1_+(R, \mathcal{A}_{sa})$ and all $T > \sup\{t \in \mathbb{R} \,|\, h_t \neq 0\}\ :\ L^h_T(\varphi) \geq 0$.

Upon having advanced this definition, Pusz and Woronowicz [1978] proved the following result:

THEOREM 51. *Let $\{\mathcal{A}, \varphi, \alpha\}$ be a C^*- dynamical system, and consider the following conditions: (I) φ is either a KMS state with respect to α for some temperature $\beta > 0$; or is a ground state; and (II) φ is passive in the sense of definition 50. Then:*

1. *Without further assumptions: $(I) \Rightarrow (II)$* .

2. *If furthermore: (i) $\mathcal{A}$ admits an action $\nu : G \rightarrow \text{Aut}(\mathcal{A})$ where G is a locally compact amenable group; (ii) ν commutes with the evolution α, i.e. $\forall\,(t,g) \in R \times G : \nu_g \circ \alpha_t = \alpha_t \circ \nu_g$; and (iii) φ is $\eta-$clustering with respect to the action of G. Then these conditions, taken together, entail $(II) \Rightarrow (I)$.*

REMARKS 52. The following remarks focus on part (2) of the theorem, i.e. the operational characterization of KMS states as passive.

1. In the passivity condition (II), φ has not been assumed to be invariant under the unperturbed evolution α ; in part (2) this property obtains as φ is proven to satisfy the KMS condition.

2. The condition that ν commutes with α is natural in view of the conclusion to be obtained: if an automorphism leaves invariant a KMS state, then it must commute with the evolution with respect to which this state is KMS.

3. Invariant means and amenable groups were introduced in subsection 3.5; see in particular definition 21 and remark 31(5).

4. Among the ancillary assumptions listed in (2), it is not even necessary to impose as a precondition that φ be $G-$invariant; this follows from the explicit assumption that it is $\eta-$clustering, i.e. (see definition 22):

$$\forall\; A\,,\, B \in \mathcal{A} \;:\; \eta^G\left(\varphi(\nu_g[A]\, B)\right) = \varphi(A)\,\varphi(B) \quad .$$

   Actually, this condition entails furthermore that φ cannot be decomposed in a convex combination of other $G-$invariant states.

5. In QSP, the natural candidate for G is the group of translations in *space*. Hence, in contrast with the stability conditions studied earlier, the assumed clustering property does not need to be with respect to time. This allows us to consider systems for which the evolution is *not* asymptotically abelian. This opening is significant when it comes to concrete modeling for the purposes of QSP: one may not wish to have to identify the weak-clustering with respect to the group $\{\nu_g \mid g \in G\}$ *and* any putative clustering with respect to the evolution $\{\alpha_t \mid t \in \mathbb{R}\}$.

6. In addition, Pusz and Woronowicz [Pusz and Woronowicz, 1978] propose an alternative route, replacing all the ancillary conditions in part (2) of the theorem by a strengthened form of passivity. Specifically, instead of considering a single dynamical system, they consider, for every positive integer, identical non-interacting copies $\{\mathcal{A}_k, \varphi_k, \alpha_k \mid k = 1, \ldots, N\}$ from which one constructs the collective dynamical system $\{\mathcal{A}^N, \varphi^N, \alpha^N\}$ where $\{\mathcal{A}^N = \otimes_{k=1}^N \mathcal{A}_k\,,\; \varphi^N = \otimes_{k=1}^N \varphi_k\,,$ and $\alpha^N = \otimes_{k=1}^N \alpha_k\,\}$. The perturbation h however is allowed to be a general element in $C^1_+(\mathbb{R}, \mathcal{A}^N)$, so that α^h is allowed not to act independently on each of the component systems. Then φ is said to be *completely passive* whenever for every positive integer N the

state φ^N is passive. Now, without further ado — i.e. without having to impose condition (2) in theorem 51 — the complete passivity of φ can be proven to be equivalent to the condition that φ satisfy the KMS condition. For QSP, the choice between the condition of complete passivity or condition (2) in the theorem, is largely a question of taste.

E. Thermodynamical stability.

To close this subsection, I wish to indicate how the concept of thermodynamical stability gives rise to yet another characterization of KMS states, this one without restriction on whether the states considered are to be extremal with respect to the KMS condition. To avoid technicalities, I present these considerations in the simplest case, namely where the system is a quantum spin-lattice and thus is described by a C^*−algebra $\mathcal{A} = \otimes_{k\in\mathbb{Z}^d}\mathcal{A}_k$ where the $\mathcal{A}_k$ are copies of a finite matrix algebra, say $\mathcal{M}(n,C)$, with n and d finite. Throughout $\Lambda\subset\mathbb{Z}^d$ denotes a connected finite subset of the lattice $\mathbb{Z}^d$; φ denotes a state on $\mathcal{A}$; φ_Λ denotes the restriction of φ to the finite matrix algebra $\mathcal{A}_\Lambda = \otimes_{k\in\Lambda}\mathcal{A}_k$; and ρ_Λ is the density matrix corresponding to φ_Λ. Furthermore it is convenient to assume here that the dynamics obtains from short-range — or possibly suitably tempered — interactions between the sites. The reader interested in how far the considerations presented below may be pursued will find a review in [Sewell, 2002]; among the original papers, let it suffice to mention for orientation purposes [Araki, 1974; Araki and Sewell, 1977; Sewell, 1977; Sewell, 1980b; Ruelle, 1968a; Robinson, 1971; Araki and Moriya, 2002].

A version of the second law of thermodynamics — compare with the equivalent form of the variational principle defined immediately after theorem 3 — defines the local free-energy relative to Λ at natural temperature $\beta = 1/kT$ as:

$$F_{\Lambda,\beta}(\varphi) = E_\Lambda(\varphi) - T\,S_\Lambda(\varphi) \quad \text{with} \quad \begin{cases} E_\Lambda(\varphi) & = \varphi_\Lambda(H_\Lambda) \\ \\ S_\Lambda(\varphi) & = -k\,\mathrm{Tr}\,\rho_\Lambda \log\,\rho_\Lambda \end{cases} .$$

Two states ψ and φ on $\mathcal{A}$ are said to satisfy the equivalence relation $\overset{\Lambda_o}{\sim}$ whenever they coincide outside the finite region Λ_o. We then write $\psi\sim\varphi$ whenever there exists Λ_o such that $\psi \overset{\Lambda_o}{\sim} \varphi$. For the quantum lattice considered here, one can then prove that the following limit exists

$$(101)\ \ \forall\,\psi\sim\varphi\ :\ \Delta F_\beta(\psi\,|\,\varphi) := \lim_{\Lambda\uparrow\mathbb{Z}^d}\ (F_{\Lambda,\beta}(\psi) - F_{\Lambda,\beta}(\varphi))\ .$$

For the order of the arguments ψ and φ in ΔF_β recall that mathematicians (and some philosophers) read from right to left, while most physicists seem to read from left to right. Thus, $\Delta F_\beta(\psi\,|\,\varphi)$, as written above, represents the increment of free-energy when passing from the state φ to any state ψ that differs from φ only in a finite region. Araki and Sewell [Araki and Sewell, 1977; Sewell, 1977] introduced the following definition and prove the following result; see also [Sewell, 1980b; Sewell, 2002].

DEFINITION 53. With $\Delta F_\beta(\varphi\,|\,\psi)$ as in (101), a state φ on $\mathcal{A}$ is said to be locally thermodynamically stable at natural temperature β whenever

$$\forall\,\psi \sim \varphi \;:\; \Delta F_\beta(\psi\,|\,\varphi)\,\geq 0 \quad .$$

Hence, to require that this stability condition be satisfied is indeed a variational principle: the free-energy of the state φ cannot be reduced by going to a state ψ that differs from φ only locally.

THEOREM 54. *For a state φ on a quantum lattice system of the type considered here, the following conditions are equivalent:*

1. *φ satisfies the KMS condition at natural temperature β;*
2. *φ is locally thermodynamically stable at natural temperature β.*

REMARKS 55.

1. This result involves *in an essential manner* the local structure of the system considered, namely that the global algebra $\mathcal{A}$ is the C^*-inductive limit of local algebras $\mathcal{A}_\Lambda$ relative to bounded regions of space, where the indexing net $\mathcal{F} := \{\Lambda\}$ is absorbing, i.e. — recall subsection 3.4, part I — for every point x in space, there is a bounded region $\Lambda \in \mathcal{F}$ such that $x \in \Lambda$. An alternative version is requiring that for every bounded region Ω of space there exists some $\Lambda \in \mathcal{F}$ such that $\Omega \subseteq \Lambda$; both versions are acceptable in axiomatic QSP.

2. Also in contrast to reservoir stability — see theorem 48 — this result is one of *internal* consistency in the sense that it establishes the equivalence of two definitions of equilibrium for the *same* system when described from two different points of view: the microscopic KMS condition and the local aspect of thermodynamics of the system considered. In particular, the argument does not involve any coupling of the system considered with any test system.

3. Extensions of the domain of validity of the theorem are desirable. In this respect, quantum spin-lattice systems with reasonably long-range interactions have been controlled. However, some technical difficulties often stand in the way toward the expected extensions to continuous systems. Typically these difficulties originate in the infinite dimensionality of the Hilbert spaces $\mathcal{H}_\Lambda$ corresponding to finite regions, and in the fact that the corresponding Hamiltonians H_Λ are unbounded; also, precautions may have to be taken to ensure that the local particle-density remains bounded.

4. One type of extension of the above variational principle is instructive, namely the shift from local stability to global stability requirements. Specifically, consider again a quantum-lattice system defined on $\mathbb{Z}^d$. Assume further that the dynamics is invariant under the translation group $G = \mathbb{Z}^d$, and restrict attention to the set $\mathfrak{S}^G$ of states ψ each of which is $G-$invariant. Assume finally that the following limits exist

(102) $f_\beta(\psi) = \lim_{\Lambda\uparrow\mathbb{Z}^d} |\Lambda|^{-1} F_{\Lambda,\beta}(\psi) \quad ; \quad \phi_\beta = \lim_{\Lambda\uparrow\mathbb{Z}^d} |\Lambda|^{-1} \log \mathrm{Tr} \exp^{-\beta H_\Lambda} \quad .$

A state $\varphi \in \mathfrak{S}^G$ is now said to be *globally thermodynamically stable* — or GTS for short — whenever it minimizes the free-energy density, i.e. when

(103) $f_\beta(\varphi) = \min_{\psi\in S_G} f_\beta(\psi) = \phi_\beta \quad .$

As long as one remains with $G-$invariant states on quantum-lattices having $G-$invariant dynamics with only *short-range* interactions, one has

$$\varphi \text{ is GTS} \iff \varphi \text{ is KMS} \quad .$$

However, while $\Rightarrow$ remains valid even when interactions are allowed to extend over a reasonably long range, the “short-range” requirement is essential for $\Leftarrow$. It has been suggested [Sewell, 1980b] that KMS states that are not GTS, i.e. do not minimize the free-energy density, may model metastable states.

5.5 A brief excursion into QFT

As a remark on the role of KMS states in mathematical physics I wish to mention, however briefly, the appearance of modular structures beyond the confines of non-relativistic QSP, namely their entry into relativistic QFT. For the general framework of algebraic QFT, cf. e.g. in this volume [Halvorson, 2006]; for a presentation specifically geared to QFT on curved space-times, cf. also [Wald, 1994]; and for a discussion of some of the interpretation problems raised by the materials in this section, cf. [Clifton and Halvorson, 2001].

From the perspective developed in this essay, the natural entry into the considerations to be discussed in the present subsection is through a manifestation, in Minkowski-space QFT, of the Tomita–Takesaki duality — recall scholium 33 or theorem 39.

Bisognano and Wichmann [Bisognano and Wichmann, 1975] developed a consequence of a standard result in axiomatic QFT — the Reeh–Schlieder theorem, cf. e.g. [Streater and Wightman, 1964, p.168], or [Emch, 1972a, p. 290] and references cited therein — which ensures in particular that the vacuum state φ, when restricted to a wedge $W_R = \{(x,y,z,t) \in M^{3+1} \mid z > |t|\}$, is faithful on the corresponding algebra $\mathcal{N}_R$. Thus, this restriction φ_R of φ to $\mathcal{N}_R$ equips the latter with the structure of a Tomita–Takesaki modular algebra. Here, the canonical objects of the Tomita–Takesaki theory have a seminal geometric interpretation. The involutive antiunitary operator J — corresponding to the reflection $(x,y,z,t) \to (x,y,-z,-t)$ which maps the wedge W_R to the wedge $W_L = \{(x,y,z,t) \in M^{3+1} \mid z < |t|\}$ — implements a bijection from $\mathcal{N}_R$ to ${\mathcal{N}_R}' \simeq \mathcal{N}_L$; and the modular group $\{\Delta^{i\lambda} \mid \lambda \in R\}$ implements on $\mathcal{N}_R$ the Lorentz boost

$$\begin{pmatrix} x \\ y \\ z \\ t \end{pmatrix} \quad \to \quad \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \cosh(2\pi\lambda) & -\sinh(2\pi\lambda) \\ 0 & 0 & -\sinh(2\pi\lambda) & \cosh(2\pi\lambda) \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ t \end{pmatrix} \quad .$$

Since uniformly accelerated observers moving in the interior of a wedge W_R perceive its boundaries as past and future horizons, the result of Bisognano and Wichman could be interpreted as saying that in the universe of such observers — the wedge W_R — the vacuum of the field in M^{3+1} appears to be a thermal bath, in the following sense. The state $\varphi_R : N \in \mathcal{N}_R \mapsto \varphi(N) \in \mathbb{C}$ — where $\mathcal{N}_R \subset \mathcal{N}$ is the algebra corresponding to the wedge W_R, $\mathcal{N}$ is the algebra corresponding to the full Minkowski space, and φ is the vacuum defined on $\mathcal{N}$ — is a KMS state at temperature $\beta > 0$ with respect to the evolution $\{\tau_t : N \in \mathcal{N}_R \mapsto \tau_t[N] = \Delta^{-it/\beta} N \Delta^{it/\beta} \in \mathcal{N}_R \mid t \in \mathbb{R}\}$ (where, as usual, the numerical value of the natural temperature $\beta = 1/kT$ depends on the scale with respect to which the time t is measured).

The physical interest of this interpretation is enhanced by an earlier remark by Rindler [Rindler, 1966] to the effect that the universe of uniformly accelerated observers in W_R is similar to the universe around the Schwarzschild solution of the Einstein equations, i.e. around a stationary "black hole".

With this dictionary in hand, the phenomenon discovered by Bisognano and Wichmann as a consequence of the Tomita–Takesaki theory translates into an effect found independently by Unruh [1976] in an attempt to clarify the then recently discovered Hawking effect [Hawking, 1975] (also known as the Hawking radiation). The latter describes a related but different phenomenon, the creation of thermally distributed particles around a collapsing black hole. The similarities and differences between the Unruh and the Hawking effects are discussed in [Wald, 1994, chapters 5 and 7]; for some of the thermodynamical aspects of the subject in the astrophysics literature, cf. e.g. [Davies, 1978; Hawking and Page, 1983] or [Wald, 1994, chapter 6]; for the specific questions of *what* is actually measurable, *how* and *where,* see [Unruh and Wald, 1984]; and for some of the philosophical issues, see [Clifton and Halvorson, 2001].

The impact of the Bisognano and Wichmann discovery on the axiomatic QFT literature began with the work of Sewell [Sewell, 1980a; Sewell, 1982a] who generalized their results to some curved manifolds, perceived the role that the bifurcate horizon plays in the Unruh effect, and proposed to identify the Hawking temperature and the temperature in the KMS condition associated with the Tomita–Takesaki modular theory. The introduction of KMS structures in QFT then turned out to be the harbinger of a *"revolution"* [Borchers, 2000]. A few among the many developments that ensued are: an extension of the axiomatic of algebraic QFT to curved manifolds; an interpretation of the intrinsic geometry of space-time in terms of consistency relations between the modular structures to be attached to an absorbing net of intersecting wedge-like regions of GR space-times; and the beginnings of a relativistic QSP where local KMS conditions are formulated in terms of future-directed time-like vectors that determine local rest-frames; cf. e.g. [Summers and Verch, 1996; Buchholz *et al.*, 2002; Ojima, 2003; Wiesbrock, 1997; Buchholz, 2003; Summers and White, 2003; Buchholz and Lechner, 2004]; closer to the Hawking effect proper, [Haag *et al.*, 1994; Kay and Wald, 1991; Fredenhagen and Haag, 1990]; and for a new framework [Fredenhagen, 2003].

5.6 A mathematical interlude: extremal KMS states

The role of extremal KMS states in QSP will be studied in subsection 5.7. The purpose of this section is to review some mathematical preliminaries such as the definition of extremal KMS states, their characterization in terms of their GNS representation, and the decomposition of a KMS state into its extremal components.

REMARKS 56.

1. Let $\mathcal{A}$ be a C^*-algebra, $\beta > 0$ and τ be a group of automorphisms of $\mathcal{A}$. The set $\mathfrak{S}_\beta$ of all KMS states on $\mathcal{A}$ that satisfy the KMS condition for τ and β is convex, i.e. for any two KMS states ψ and χ on $\mathcal{A}$, with respect to the same τ and β, and any $\lambda \in (0,1)$: $\varphi = \lambda\psi + (1-\lambda)\chi$ is again a KMS state for τ and β.

2. The set $\mathfrak{S}_\beta$ is closed in the w^*- topology it inherits from $\mathcal{A}$, and it is bounded in the metric topology. Hence it is w^*-compact, and the Krein–Milman theorem entails that $\mathfrak{S}_\beta$ is the w^*-closed convex hull of the set $\mathfrak{E}_\beta$ of its extreme points [Dunford and Schwartz, 1964, theorem V.8.4]. This ensures not only the existence of extremal points, but also that there are sufficiently many of them: every element in $\mathfrak{S}_\beta$ is the limit of finite convex sums of elements in $\mathfrak{E}_\beta$; see definition 57 below.

3. Moreover $\beta_1 \neq \beta_2$ entails $\mathfrak{S}_{\beta_1} \cap \mathfrak{S}_{\beta_2} = \emptyset$. Incidentally, the GNS representations constructed from states $\varphi_1 \in \mathfrak{S}_{\beta_1}$ and $\varphi_2 \in \mathfrak{S}_{\beta_2}$ with $\beta_1 \neq \beta_2$ are disjoint in the sense that no subrepresentation of one of these is unitarily equivalent to any subrepresentation of the other; cf. [Takesaki, 1970c].

DEFINITION 57. Given a von Neumann algebra $\mathcal{N}$, a group $\{\tau_t \mid t \in \mathbb{R}\}$ of automorphisms of $\mathcal{N}$, $\beta \in \mathbb{R}^+$, and $\mathfrak{S}_\beta$ as in remark 56(1) above. A state $\varphi \in \mathfrak{S}_\beta$ is said to be extremal KMS at natural temperature β if it does not admit a convex decomposition into states in $\mathfrak{S}_\beta$ — i.e. states that satisfy the KMS condition for the same τ and β. The set of all extremal KMS states is denoted $\mathfrak{E}_\beta$.

THEOREM 58. *Let φ be a faithful normal state on a von Neumann algebra $\mathcal{N}$ and τ be the unique group of automorphisms of $\mathcal{N}$ with respect to which φ satisfies the KMS condition for some natural temperature β. Denote by $\mathcal{Z}$ the center $\mathcal{N} \cap \mathcal{N}'$ of $\mathcal{N}$. Then*

A. *For every* $(t, Z) \in \mathbb{R} \times \mathcal{Z}$, $\tau_t[Z] = Z$.

B. *For every positive non-zero element* $Z \in \mathcal{Z}$ *with* $0 < Z < I$,

$$\psi(N) := \varphi(Z)^{-1}\varphi(ZN) \quad \text{and} \quad \chi(N) := \varphi(I-Z)^{-1}\varphi((I-Z)N)$$

define two states ψ *and* χ *on* $\mathcal{N}$ *that satisfy the KMS condition for the same* τ *and* β *and provide a convex decomposition of* φ .

C. *For every* φ *that admits a convex decomposition* $\varphi = \lambda\psi + (1-\lambda)\chi$ *into states* ψ *and* χ *on* $\mathcal{N}$ *that satisfy the KMS condition for the same* τ *and* β , *there exists a unique positive non-zero element* $Z \in \mathcal{Z}$ *with* $\|Z\| \leq 1$ *such that for all* $N \in \mathcal{N}$

$$\psi(N) = \varphi(Z)^{-1}\varphi(ZN) \quad \text{and} \quad \chi(N) = \varphi(I-Z)^{-1}\varphi((I-Z)N) .$$

Proof. As pointed out in remark 37, we may assume without loss of generality that $\mathcal{N}$ is presented in standard form, so that there exists a cyclic and separating vector $\Phi \in \mathcal{H}$ for $\mathcal{N}$ with $\forall N \in \mathcal{N} : (\Phi, N\Phi) = \varphi(N)$.

[A.] $Z \in \mathcal{Z} \Rightarrow \forall (t, N) \in \mathbb{R} \times \mathcal{N}$, $\varphi(N^*\tau_t[Z]) = \varphi(\tau_t[z]N^*)$ and thus φ being KMS entails that $\varphi(N\tau_t[Z])$ is constant in t so that $\forall t \in \mathbb{R} : (N\Phi, [\tau_t[Z] - Z]\Phi) = 0$. Φ being cyclic entails $[\tau_t[Z] - Z]\Phi = 0$, and then Φ being separating entails $[\tau_t[Z] - Z] = 0$.

[B.] φ being faithful and $0 < Z < I$ positive and non-zero entail $0 < \varphi(Z) < 1$; and, upon taking into account that Z and thus $Z^{\frac{1}{2}}$ belong to $\mathcal{N}'$, one verifies that ψ and χ are states on $\mathcal{N}$ and that they inherit from φ its KMS property. Moreover, one reads immediately from their definition that $\varphi = \lambda\psi(N) + (1-\lambda)\chi(N)$, where $0 < \lambda = \varphi(Z) < 1$.

[C.] Conversely, from $\varphi = \lambda\psi(N) + (1-\lambda)\chi(N)$, with $0 < \lambda < 1$ one has $\psi \leq \lambda^{-1}\varphi$ and thus there exists an element $X \in \mathcal{N}'$ such $\forall N \in \mathcal{N} : \psi(N) = (X\Phi, NX\Phi)$, i.e. ψ is a vector state on $\mathcal{N}$ and thus is normal and majorized by the normal functional $\lambda^{-1}\varphi$. Hence the Sakai-Radon-Nikodym [Sakai, 1971, proposition 1.24.4], entails that there exists some positive $Y \in \mathcal{N}$ with $\|Y\| \leq 1$ such that

$$\forall N \in \mathcal{N} : \psi(N) = \frac{1}{2}\lambda^{-1}\varphi(NY + YN) .$$

Suppose that there exists another element $\tilde{Y} \in \mathcal{N}$ with the same properties. Let then $X = Y - \tilde{Y}$. We have then $0 = \varphi(X^*X + XX^*)$ and thus, since φ is a positive linear functional and both X^*X and XX^* are positive: $\varphi(X^*X) = 0$. Since φ is faithful, $X = 0$ i.e. $Y = \tilde{Y}$ i.e. Y is unique.

It remains to be shown that the assumptions of the theorem entail that Y also belongs to $\mathcal{N}'$. Since φ and ψ are KMS, they satisfy for all $t \in \mathbb{R}$: $\varphi \circ \tau_t = \varphi$ and $\psi \circ \tau_t = \psi$. Consequently

$$\psi(N) = \psi(\tau_t[N]) = \frac{1}{2}\lambda^{-1}\phi(\tau_t[N]Y + Y\tau_t[N]) = \frac{1}{2}\lambda^{-1}\phi(N\tau_{-t}[Y] + \tau_{-t}[Y]N) .$$

From the uniqueness of $Y \in \mathcal{N}$ which we just established, we have $\forall t \in \mathbb{R}$: $\tau_t[Y] = Y$. φ being KMS entails therefore $\forall N \in \mathcal{N} : \varphi(NY) = \varphi(YN)$ and thus $\psi(N) = \lambda^{-1}\varphi(YN)$. Upon applying the KMS condition to both ψ and φ, we get $\forall N \in \mathcal{N} : NY = YN$ i.e. $Y \in \mathcal{N}'$. Clearly then $\lambda = \varphi(Z)$. The same argument goes through with χ replacing ψ and $(I - Z)$ replacing Z. ■

The following characterization is an immediate consequence of the above theorem:

COROLLARY 59. *With the assumptions of theorem 58, the KMS state φ is extremal KMS iff $\mathcal{N}$ is a factor, i.e. iff $\mathcal{N}$ has trivial center: $\mathcal{N} \cap \mathcal{N}' = \mathbb{C}I$.*

SCHOLIUM 60. With the assumptions of theorem 58, assume that φ is not extremal KMS, but that the center $\mathcal{Z}$ of $\mathcal{N}$ is generated by a family $\{P_k \in \mathcal{Z} \mid k = 1, 2, ...\}$ of mutually orthogonal projectors. Then there exists a unique decomposition of φ into a convex combination $\sum_k \lambda_k \varphi_k$ of states φ_k on $\mathcal{N}$ where the φ_k are extremal KMS for the same dynamics τ and the same natural temperature β.

Proof. To say that φ is a KMS state that is not extremal KMS is to say that there exist KMS states ψ_j and scalars $\mu_j \in (0,1)$ such that $\varphi = \sum_j \mu_j \psi_j$. From part C of the theorem, for every ψ_j there exists a positive $Z_j \in \mathcal{Z}$ such that $\forall N \in \mathcal{N} : \psi_j = \phi(Z_j)^{-1}\phi(Z_j N)$. Since $\mathcal{Z}$ is an abelian von Neumann algebra with discrete spectrum, every Z_j may be written as $\sum_k z_k P_k$ with $z_k \in \mathbb{R}^+$ and the P_k are minimal projectors in $\mathcal{Z}$. Hence the $\varphi_k : N \in \mathcal{N} \mapsto {\lambda_k}^{-1}\, \varphi(P_k N) \in \mathbb{C}$ with $\lambda_k = \phi(P_k)$ are states on $\mathcal{N}$. From part B of the theorem, these are still KMS states for the same τ and β. Therefore, it only remains to show that the states φ_k are extremal with respect to the KMS condition.

To see this, consider the decomposition $\mathcal{H} = \oplus_k \mathcal{H}_k$ where $\mathcal{H}_k$ are the subspaces $\{\Psi_H \in \mathcal{H} \mid P_k\Psi = \Psi\}$. Since each P_k belongs to $\mathcal{Z}$, the subspaces $\mathcal{H}_k$ are stable under $\mathcal{N}$ and under $\mathcal{N}'$, i.e. whenever $X \in \mathcal{N}$ or $X \in \mathcal{N}'$, we have $\forall \Psi \in \mathcal{H}_k : X\Psi \in \mathcal{H}_k$. Let then $\mathcal{N}_k = \{P_k N P_k \mid N \in \mathcal{N}\}$, $\mathcal{N}'_k = \{P_k N P_k \mid N \in \mathcal{N}'\}$; and note that these are von Neumann algebras acting on the space $\mathcal{H}_k$ admitting there a cyclic and separating vector, namely $P_k\Phi$, such that $\forall N \in \mathcal{N}_k$: $\tilde{\varphi}_k(N) := (\Phi_k, N\Phi_k)$ defines a faithful normal state on $\mathcal{N}_k$; it is thus the restriction to this algebra of the state φ. Note further that for all $t \in \mathbb{R}$, $\mathcal{N}_k$ is stable under τ_t. Since $\mathcal{N}_k \cap {\mathcal{N}_k}' = \mathbb{C}I_k$ (where I_k is the identity operator in $\mathcal{H}_k$) $\tilde{\varphi}_k$ is extremal KMS. Proceeding *ab absurdo*, suppose that φ_k itself is not extremal KMS. Then there would exist some KMS state ψ on $\mathcal{N}$ and some $\lambda \in (0,1)$ such that $\psi \leq \lambda^{-1}\varphi_k$. Denote by $\tilde{\psi}_k$ the restriction of ψ to $\mathcal{N}_k$. We have then, in particular, $\lambda^{-1}\varphi_k(N^*N) \geq \psi([NP_k]^*[NP_k]) = \psi_k(N^*N)$; i.e. $\lambda^{-1}\tilde{\varphi} \geq \tilde{\psi}_k$. Since $\tilde{\varphi}_k$ is extremal KMS and $\tilde{\psi}_k$ is KMS, the equality must prevail, i.e. $\lambda^{-1}\tilde{\varphi} = \tilde{\psi}_k$; and since $\tilde{\varphi}_k$ and $\tilde{\psi}_k$ are states, $\lambda = 1$, i.e. on $(N)_k : \tilde{\psi}_k(N^*N) = \tilde{\varphi}_k(N^*N)$. By the Schwartz inequality, we have for every $N \in \mathcal{N}$, $\tilde{\psi}_k([P_k N P_k]^*[P_k N P_k]) \leq \psi(N^*N)$, and thus $\psi \geq \varphi_k$. Together with the initial inequality, namely $\psi \leq \varphi_k$ (since we know now that $\lambda = 1$), these two inequalities reduce to $\psi = \varphi_k$. Hence φ_k is indeed an extremal KMS state on $\mathcal{N}$. Since φ_k is extremal KMS on $\mathcal{N}_k$, the restriction

$\tilde{\psi}_k$ of ψ to this algebra must coincide with $\tilde{\varphi}_k$; and thus φ_k is maximal KMS on $\mathcal{N}$. Hence φ has been decomposed into a convex combination of extremal KMS states. Uniqueness follows by contradiction. ■

DEFINITION 61. A convex set $\mathcal{C}$ is said to be a simplex whenever every point in $\mathcal{C}$ admits a unique convex decomposition into extremal points of $\mathcal{C}$.

Recall that in two-dimensional Euclidean geometry, a triangle is a simplex; indeed any point in the triangle obtains as a unique convex combination of points situated at the vertices of the triangle. But a circle is not a simplex: the set of its extreme points is the circumference of the circle, and given any point inside the circle, all secants through this point give different convex combinations of extreme points.

REMARKS 62.

1. Scholium 60 may therefore be paraphrased by saying that under the assumption that the spectrum $Sp(\mathcal{Z})$ of the center $\mathcal{Z}$ is discrete, $\mathfrak{S}_\beta$ is a simplex; and that the decomposition is a weighed sum with respect to a discrete probability measure supported by the extreme points $\mathfrak{E}_\beta$ of the set $\mathfrak{S}_\beta$ of all normal KMS states for the given dynamics τ and the given natural temperature β . From the proof of the scholium, one checks that the latter statement extends indeed to all normal states, and not just to those that are faithful.

2. In case $Sp(\mathcal{Z})$ is not discrete, the above sum must be replaced by an integral, and some measure-theoretical trimmings are necessary to specify the sense in which φ defines a *unique* measure concentrated on the boundary of $\mathfrak{S}_\beta$. The general mathematical context in which these decompositions appear is in the study of central measures, see [Takesaki, 1970a; Kadison and Ringrose, 1983/1986]. For the purpose of this essay, the simpler version just described will suffice to anchor the conceptual structure of the theory governing the unique decomposition of KMS states into their extremal components.

3. Note that the set of states on a quantum system described by the von Neumann postulates is *not* a simplex: if a density matrix has at least one eigenvalue with multiplicity greater or equal to 2, its decomposition in pure states is not unique. Hence the set of KMS states in *quantum* statistical physics possesses a *classical* property that is otherwise not heard of in the quantum realm.

4. It remains to demonstrate that this property is relevant to QSP; and thus that QSP requires the consideration of situations for which the relevant representations do not lead to factors, in contrast with the von Neumann formalism of quantum mechanics where the canonical equilibrium states lead only to factor representations — recall the end of remark 34. This problem is the object of the next subsection.

5.7 Extremal KMS states, pure thermodynamical phases

The main argument one can advance to justify the claim that pure thermodynamical phases be described in QSP as extremal KMS states originates in the conjunction of three circumstances.

The *first* is based on subsections 5.3 and 5.4 where strong evidences were given for the identification of canonical equilibrium states as KMS states.

The *second* is the fact that extremal KMS states are the elementary objects in the KMS theory. This is reminiscent of the identification of atomic levels in *spectroscopy* with irreducible representations of the group of symmetries of the system, the famous "Gruppenpest" epitomized in [Wigner, 1931]. In mathematics, this programme was extended to a systematic presentation of the familiar so-called *special functions* where these functions now appear as bases of irreducible representations of groups; cf. e.g. [Talman, 1968; Vilenkin, 1968]. Closer to the focus of this essay, the early identification — in [Murray and von Neumann, 1936] — of *factors as the building blocks* of the theory of von Neumann algebras proceeds from the same principle: a methodological option confirmed by the central decomposition of a von Neumann algebra as a direct integral of factors; cf. e.g. [Kadison and Ringrose, 1983/1986, theorem 14.2.2, pp. 1027–1028]. All the while, the group-theoretical approach continues to contribute in sorting out qualitative clssification problems in *nuclear spectroscopy,* and *elementary particles* high energy physics.

The *third* circumstance pointing to the description of pure thermodynamical phases as extremal KMS states — i.e. KMS states the GNS representations of which are factors, cf. corollary 59 above — is the mathematical fact that the decomposition of a KMS state in extremal KMS states is *unique*; cf. scholium 60 and remark 62(2) above. In the context of QSP, this fact naturally directs attention to the situation encountered in thermodynamics where an equilibrium state decomposes uniquely into its pure thermodynamical phases.

Thus, this subsection is divided into two parts. In part **A**, the above speculations are confronted with a model for QSP where everything can be computed explicitly. In part **B**, the characterization of pure thermodynamical phases as extremal KMS states is brought to bear on a famous argument by Landau offering a fundamental microscopic distinction between solids and fluids in term of space correlation functions. This exemplifies how the unique decomposition of KMS states into their extremal components helps describe the coexistence of pure thermodynamical phases in QSP and spontaneous symmetry breaking in systems undergoing phase transitions. For further discussion of the latter, cf. [Liu and Emch, 2005].

A. Quantum Weiss–Ising models for ferromagnetism

Recall first the results of Araki reported in subsection 5.3 concerning the absence of a ferromagnetic phase transition: the unique KMS state for each of the models covered there is extremal.

To check how this coincidence fares with systems that *do* exhibit several thermodynamical phases, we turn now to a class of models that have a long history in the physics of phase transitions [Weiss, 1907; Brout, 1965], and are accepted by mathematicians to be amenable to a sufficiently rigorous treatment [Kac, 1968], namely the Weiss–Ising models for ferromagnetism.

Consider a one-dimensional lattice $\mathbb{Z}$ where to every site $k \in \mathbb{Z}$ a quantum spin σ_k is attached. To every finite string $\Lambda \subset \mathbb{Z}$ is associated a Hamiltonian

$$(104)\quad H = -\sum_{k\in\Lambda}[B + B_{\Lambda,k}]{\sigma_k}^z \quad \text{with} \quad B_{\Lambda,k} = \frac{1}{2}\sum_{j\in\Lambda} J_{\Lambda,jk}{\sigma_j}^z$$

where B is interpreted as a homogeneous external magnetic field parallel to a fixed direction z; and $B_{\Lambda,k}$ is an average magnetic field, the so-called "molecular" field, experienced by the spin at site k resulting from all other spins in the region Λ. The artificial assumption imposed on the models of the van der Waals or Weiss-type, which makes them exactly solvable in the thermodynamical limit, is that the strength of the interaction $J_{\Lambda,jk}$ decreases with the size $|\Lambda|$ of Λ; compare this with the property $|v(p,q)| \leq c/|\Lambda|$ of the interaction of the BCS model in subsection 5.1.

Adopting here a simplified version of [Emch and Knops, 1970], we will assume that

$$(105)\quad J_{\Lambda,jk} = \begin{cases} |\Lambda|^{-1}\, J > 0 & \text{when} \quad j \neq k \\ 0 & \text{otherwise} \end{cases} .$$

Upon controlling the thermodynamical limit $|\Lambda| \to \infty$, one finds that two extremal KMS states emerge when $T < T_c$ where $1/kT_c = \beta_c = J^{-1}$. These are recognized by the following properties of a global observable — cf. definition 15 and scholium 23 — namely, the magnetization $\mathbf{M}$, the three components of which

$$M^i = \text{weak op. limit}_{\,|\Lambda|\to\infty}\ \frac{1}{|\Lambda|}\sum_{k\in\Lambda}{\sigma_k}^i \qquad (i = x, y, z)$$

are defined in the corresponding temperature-dependent representation. They satisfy

$$(106)\quad \begin{array}{ll} \text{(i)} & M^x = M^y = 0 \\ \text{(ii)} & M^z = \tanh[\beta(B + JM^z)]\ . \end{array}$$

REMARKS 63.

1. For the transverse components, M^x and M^y, (106.i) was to be expected from the symmetry of the system. The interesting part is the result for the component M^z parallel to the applied magnetic field: (106.ii) is the classical self-consistency equation: the model exhibits a phase transition as there exists a temperature T_c (with $\beta_c = J^{-1}$) below which M^z does not vanish as $B \to 0$, but tends to a finite, temperature-dependent value, the so-called spontaneous magnetization.

2. Hence in the thermodynamical limit, the problem of determining the extremal KMS states of the model admits two new solutions, absent above T_c; these extremal KMS states exhibit the two opposite spontaneous magnetizations characteristic of the two pure thermodynamical phases familiar from the treatment of the classical case in e.g. [Kac, 1968].

3. A phase transition has occurred at $T = T_c$ and it is accompanied, for $T < T_c$, by a spontaneous breakdown of the flip-flop symmetry ${\sigma_k}^z \to -{\sigma_k}^z$ of the local Hamiltonians (104).

4. One ought to note at this point that the treatment in [Kac, 1968] was in the most orthodox spirit of classical statistical mechanics: an analysis by steepest descent methods of the partition function in the limit $|\Lambda| \to \infty$. The novelty in [Emch and Knops, 1970] was to consider also the evolution of the $x-$ and $y-$ components of the quantum spins and to study the resulting quantum dynamics in order to confront the interpretation of extremal KMS states with results already known from a classical treatment.

   As with the BCS model reviewed in subsection 5.1 above, which has also the structure of a 'molecular' field model, some technicalities are involved here: in the thermodynamical limit, the convergence of the evolution is established only for the von Neumann algebras belonging to the representations considered.

5. This simplified version of the model, where (105) is assumed, already allows us to demonstrate the general features explored in this subsection. We may nevertheless mention that in [Emch and Knops, 1970] $J_{\Lambda,jk}$ is allowed to depend on the distance $|j-k|$, but only in such a way that for each $k \in \mathbb{Z}$, there exists a constant c_k such that $\sum_k |J_{\Lambda,jk}| < c_k$ for every finite Λ containing k, subject to the condition that $\forall\, j,k \in \mathbb{Z} \,:\, \lim_{|\Lambda|\to\infty} J_{\Lambda,jk} = 0$; the set of thermodynamical phases then becomes more complex, but its description still illustrates the adequacy of the decomposition account of spontaneous symmetry breakdown.

6. We already pointed out — first in section 4.1 — that as a consequence of the Liouville theorem of complex analysis, KMS states are necessarily time-invariant. Nevertheless, as established in [Emch and Knops, 1970], the present model admits *extremal KMS states that are not extremal time-invariant* — i.e. KMS states that are convex sums of different time-invariant, but not KMS, states — although these extremal KMS states do satisfy a very strong clustering property with respect to *space-translations.* The occurrence of extremal KMS states that are not extremal time-invariant reflects the fact that the time-evolution is *not* asymptotically abelian. This is not an unexpected peculiarity of the model since experience shows that asymptotic abelianness for the group of *time*-translations is rarely satisfied in QSP — although a few exceptions are known, among them the even part of the XY-model discussed at the beginning of subsection 5.4, despite the fact that

locality entails very strong asymptotic abelianness for the group of *space*-translations.

The coexistence of liquid and gas — say vapour and liquid water — presents formal similarities with the coexistence of magnetic phases oriented in opposite directions. The lattice-gas models of classical statistical mechanics are treated in close analogy with those of their ferromagnetic counterparts: instead of attributing to each site of a regular n-dimensional Ising model, a classical spin taking the values $+1/2$ and $-1/2$, one considers a random variable indexed by the sites of the lattice and taking the values 1 or 0 depending on whether the site is occupied by a molecule or not; double (or higher) occupancy is ruled out by *fiat* in these models. Phenomenologically, liquid-vapour coexistence curves in the phase diagram translate closely to the coexistence curves in ferromagnetic materials. In particular both present a critical point, precisely located in the phase space by the occurrence of diverging fluctuations. For temperatures higher than the critical temperature, any distinction between liquid and gas is untenable, and this state of matter is best described as a fluid.

B. QSP brought to bear on the Landau argument

The situation encountered with the coexistence of fluid and crytalline phases of the same substance — say water in its fluid phase and ice phase — is phenomenologically very different from the situation presented by a gas-liquid phase transition. Here, no critical point has been located: the fluid-solid coexistence curve extends indefinitely as pressure and density are increased. A heuristic argument for the non-existence of a critical point for the fluid-solid coexistence curve was advanced by Landau; see for instance [Landau and Lifshitz, 1958b, p. 260]. The argument was taken up by Uhlenbeck in [Uhlenbeck, 1968, p. 17]: "Because the solid and the fluid are with respect to long range order qualitatively different, there cannot be a critical point, since by going around it this would imply that long range order would appear gradually, which is impossible. This is the argument of Landau and I find it completely convincing." And yet, Uhlenbeck warns on the same page that "one cannot escape the fact (intuitively evident, although not proved!) that there is already long range order in the solid phase itself."

In an impressive sequence of papers, Kastler *et al.* [1967] rose to the challenge; the various assumptions of asymptotic abelianness, pervasive in these papers, was shown to be dispensable in the version worked out in [Emch *et al.*, 1970], which is followed here.

The programme is to classify the extremal KMS states that appear in the decomposition of a Euclidean invariant KMS state. Let $\mathcal{A}$ be the C^*-algebra obtained as the C^*-inductive limit of local algebras $\mathcal{A}(\Lambda)$ over an absorbing net $\mathcal{F}$ of finite regions $\Lambda \subset \mathbb{R}^3$ (here 'finite' means finite volume: $|\Lambda| < \infty$). Let $\alpha : t \in \mathbb{R} \mapsto \alpha_t \in \mathrm{Aut}(\mathcal{A})$ describe an evolution; let $\nu : g \in \mathbb{E}^3 \mapsto \alpha_g \in \mathrm{Aut}(\mathcal{A})$ describe the action of the Euclidean group $\mathbb{E}^3$; and let φ be a KMS state on $\mathcal{A}$ with respect to the evolution α for the temperature β ; φ is assumed to be invariant under the action of the Euclidean group, i.e. $\forall g \in \mathbb{E}^3 \,:\, \varphi \circ \nu_g = \varphi$; this condition is motivated by the phenomenological expectation that the underlying interactions are Euclidean invariant.

It is convenient to assume further that φ is strongly transitive with respect to the action of $\mathbb{E}^3$ in the sense that the following two conditions are satisfied.

1. For any two states ψ and ψ' appearing in the decomposition of φ in extremal KMS states, there exists at least one $g \in \mathbb{E}^3$ such that $\psi' = \psi \circ \nu_g$.

2. For one — and therefore all — state ψ appearing in the decomposition of φ into extremal KMS states, the isotropy subgroup $G_\psi := \{g \in \mathbb{E}^3 \mid \psi \circ \nu_g = \psi\}$ contains at least three non-coplanar translations.

Note that for any $g \in \mathbb{E}^3$ and any ψ appearing in the decomposition of φ into extremal KMS states, the state $\psi_g := \psi \circ \nu_g$ also appears there; and that $G_{\psi_g} = g^{-1} G_\psi g$. Hence, up to conjugacy, all elements appearing in the decomposition of φ have the same symmetry. This conjugacy class is denoted G^φ , and is referred to it as the *intrinsic symmetry* of φ . It is the part of the Euclidean symmetry of φ that is preserved when φ is decomposed into its extremal KMS components. Consequently, condition (1) is essentially one of convenience: if it were not satisfied, one would first have to separate the decomposed states in classes of conjugate elements, and carry out the analysis sketched below for each class separately. Condition (2) excludes pathological cases which one does not want to consider here. Mathematically, it strengthens condition (1) to ensure that the orbit of each extremal state under the translation group $\mathbb{R}^3 \subset \mathbb{E}^3$ in the space of all states on $\mathcal{A}$ is closed.

It is then proven in [Emch *et al.*, 1970] that a Euclidean-invariant KMS state φ that satisfies the above conditions must necessarily belong to one of the following four classes.

<u>The first class</u> obtains when φ is already extremal KMS, i.e. its intrinsic symmetry is the group $\mathbb{E}^3$ itself. This case occurs *exactly when* one — and thus all — of the following *equivalent* conditions is satisfied:

1. φ is extremal $\mathbb{R}^3$ invariant, i.e. cannot be decomposed into a convex combination of states that are invariant under all translations in $\mathbb{R}^3$.

2. The spectrum of the generator $\mathbf{P}$ of the unitary representation of $\mathbb{R}^3$ canonically associated to φ by the GNS construction consists of exactly one eigenvalue, namely $\mathbf{k} = 0$, and this eigenvalue is non-degenerate.

3. φ is uniformly clustering in space, i.e. : for every $\epsilon > 0$ and $A \in \mathcal{A}$ there exists a finite region of space $\Lambda \subset \mathbb{R}^3$ such that

   (107) $\forall\, B \in \mathcal{A}(\Lambda^c) : |\varphi(AB) - \varphi(A)\varphi(B)| \leq \epsilon \|B\|$

   where $\mathcal{A}(\Lambda^c) \subset \mathcal{A}$ is the C^*-inductive limit of the local algebras $\mathcal{A}(\Omega)$ with $\Omega \in \mathcal{F}$ and $\Omega \bowtie \Lambda$, (i.e. $\Omega \cap \Lambda = \emptyset$); see definition 27, scholium 29 and corollary 30 above.

In view of these properties, a state φ belonging to this class is intepreted as a *fluid phase.*

To describe the other three classes, namely the strongly transitive Euclidean invariant KMS states that do *not* describe fluids, let us focus now on the notion of the *intrinsic translational invariance* of φ. For any state ψ that appears in the decomposition of φ into extremal KMS states, let G_ψ denote the subgroup of Euclidean symmetries of ψ, and let $H_\psi = G_\psi \cap \mathbb{R}^3$ denote the subgroup of space-translations that preserve ψ. As one reviews the definition of the conjugacy classes one verifies that this group is indeed characteristic of the original state φ. Note also that strong transitivity entails that $\mathbb{R}^3/H_\psi$ is compact.

The second class of Euclidean, strongly transitive KMS states is now specified by the following *equivalent* conditions, where ψ is any state appearing in the decomposition of φ into its extremal KMS components.

1. G_ψ is a crystallographic group.

2. φ is not extremal $\mathbb{R}^3-$invariant, and H_ψ is generated by three non-coplanar translations.

3. With $\chi = \eta^{\mathbb{R}^3}[\psi]$ — where $\eta^{\mathbb{R}^3}$ is any invariant mean over the translation group $\mathbb{R}^3$ — χ is $\eta-$clustering (see definition 22 above), but neither weakly mixing nor even partially weakly mixing, i.e. χ satisfies

   $$(108)\ \ \forall\, A\,,\, B \in \mathcal{A} \,:\ \eta^{\mathbb{R}^3}\left(\chi(\nu_\bullet[A]\,B) - \chi(A)\,\chi(B)\right) = 0$$

   but does *not* satisfy any of the stronger conditions

   $$(109)\ \ \forall\, A\,,\, B \in \mathcal{A} \,:\ \eta^{\mathbb{R}^3}\,|\,\chi(\nu_\bullet[A]\,B) - \chi(A)\,\chi(B)\,| = 0$$

   $$(110)\ \ \forall\, A\,,\, B \in \mathcal{A} \,:\ \eta^{\mathbb{R}^1}\left|\,\eta^{\mathbb{R}^2}\left(\chi(\nu_\bullet[A]\,B)\right) - \chi(A)\,\chi(B)\right| = 0$$

   $$(111)\ \ \forall\, A\,,\, B \in \mathcal{A} \,:\ \eta^{\mathbb{R}^2}\left|\,\eta^{\mathbb{R}^1}\left(\chi(\nu_\bullet[A]\,B)\right) - \chi(A)\,\chi(B)\right| = 0$$

REMARKS 64.

1. Taken separately, each of the conditions (1–3) excludes that φ be a fluid phase. Indeed, a fluid phase is extremal KMS, so that its intrinsic symmetry is the Euclidian group $\mathbb{E}^3$, contradicting (1); a fluid phase is extremal $\mathbb{R}^3-$invariant, contradicting (2); a fluid phase is uniformly clustering (see 5.33), which implies that each of the relations (108–111) would be satisfied, whereas (109–111) are not satisfied in the present phase.

2. The other two classes to which φ may belong are characterized as follows. (109), if satisfied, would have entailed $H_\psi = \mathbb{R}^3$, thus contradicting the second part of condition (2). This would correspond to a situation where the rotational symmetry is broken whereas the translational symmetry of the state φ would be completely preserved in its decomposition into its extremal KMS components. Although this may occur in systems exhibiting spontaneous magnetization, it is not immediately relevant to the purpose of identifying the way in which the formalism distinguishes fluids from solids in a world where fundamental interactions are invariant under the Euclidean group $\mathbb{E}^3$.

   Similarly, (110) or (111), if satisfied, would have entailed H_ψ is continuous in one or two direction(s) but discrete in the complementary direction(s). Such situations have been envisaged also — as early as the mid 1930s, cf. e.g. [Landau and Lifshitz, 1958b, p. 410] — but here again, their putative existence does not bear directly on the problem at hand.

3. The space-averaged state χ, which is $\mathbb{R}^3-$invariant by construction, nevertheless keeps a memory of the symmetry of the state ψ from which it is constructed. Indeed the *discrete* part of the spectrum of the generator $\mathbf{P}_\chi$ of the unitary group representation of $\mathbb{R}^3$, associated to the GNS construction corresponding to χ, coincides with the reciprocal group of H_ψ, namely with

   (112) $H_\psi{}^* = \{\mathbf{k} \in \mathbb{R}^3 \mid \forall \mathbf{a} \in H_\psi \,:\, \mathbf{k}\cdot\mathbf{a} = 0 \mod 2\pi\}$

   which, in principle, is observable in X-ray diffraction patterns.

Upon keeping from the above what is relevant to the absence of a critical point in the coexistence curve between fluid and solid, the analysis of the decomposition of Euclidean invariant canonical equilibrium states into their pure thermodynamical phase components provides a rigid relation between the clustering properties and the geometric properties of these states. Namely: fluid phases exhibit a uniform clustering property (107), while crystalline phases lead to the distinctly weaker property of weak-clustering (108), thus vindicating the Landau argument.

6 WHENCE AND WHITHER QSP?

This final section may serve as a summary, a conclusion, a collection of appendices sharpening some aspects of the theory that have not been discussed in the main text; and hopefully, as a constructive prospectus for territories beyond the scope of this essay.

Let me summarize very briefly the story so far. First, recall that sections 1 to 3 reviewed some of the salient features that are variously treated in traditional texts on QSP. Then, in sections 4 and 5, I argued that the algebraic formalism of the KMS condition provides a well-defined syntax, the semantics of which supports the following associations in equilibrium QSP.

- Canonical equilibrium states are described by KMS states, a notion that translates naturally from finite systems to systems considered in the themodynamical limit.
- Pure thermodynamical phases are described by extremal KMS states.
- Pure thermodynamical phase components that appear in the unique decomposition of a canonical equilibrium state may have a symmetry lower than that of the original state: only the manifold of the different contributing phases reflects the original symmetry; cf. [Liu and Emch, 2005] where we describe the "decompositional account" of spontaneous symmetry breakdown in the quantum theory of phase transitions.

Against this background, the material of the present section is presented in four subsections. I first review the mathematical concept of a limit and its physical interpretation as used in the main text. I then discuss again the notion of macroscopic observables, taking here a perspective that opens on the next subsection: the quantum measurement problem. Finally, I present some remarks — prospective and/or revisionist? — on the pursuit of constructive confrontations between mathematical and theoretical physicists in order that they better inform the wider arenas where philosophers of science operate.

6.1 Four limiting procedures in QSP

In dealing with the topics just reviewed, and as early as in sections 2 and 3, at least four different types of limits were encountered, alone or in concert.

1. the classical limit $h \to 0$;
2. the high temperature limit $T \to \infty$;
3. the thermodynamic limit $|\Lambda| \to \infty$;
4. the van Hove limit $\{\lambda \to 0 \text{ and } t \to \infty\}$ with $\tau := \lambda^2 t$ remaining finite.

As the philosophical legitimacy of each of these four limits (or 'limiting procedures') has been variously questioned elsewhere, I ought to specify again — in the vernacular, i.e. without an explicit mention of the traditional $(\epsilon, \delta_\epsilon)$ — that the limits were consistently understood in this essay to be *controlled limits* in the sense of mathematics: you give me a tolerance, and I tell you the price; the smaller the tolerance, the higher the price; but however small the error you are willing to tolerate, there is a price under which you are guaranteed that the article will be within what you decided you are going to tolerate. Mathematical physics adds to this the requirement that the "price" be expressed in currencies recognized by the putative laboratory technician. Let us examine successively the above four limits from this perspective.

1. The classical limit.

The Planck constant is a *fundamental physical constant:* in cgs units $h \simeq 6.62 \times 10^{-27}$ ergsec; the familiar notation $\hbar := h/2\pi$ is used here. To say that it is small is a "value judgement", reflecting the energy scale which you believe is relevant for the problem you wish to discuss. To illustrate the working of limiting processes, and their physical meaning, let us examine a specific example, the classical limit of a typically quantum phenomenon, the tunnel effect in which a particle of energy E does "slip through" a barrier of height $V_o > E$. This effect was discovered in 1928 independently by Gamow and by Gurney and Condon [Gamov, 1928; Gurney and Condon, 1928; Gurney and Condon, 1929] in their search for an explanation of alpha-particle emission from heavy nuclei. The Josephson junction — an oxide layer sandwiched between two superconductors — is a more recent manifestation of this quantum phenomenon; cf. e.g. [Josephson, 1982]. Let us consider here the simplest model, quantum tunelling through a square one-dimensional barrier.

One verifies immediately that the Schrödinger equation

$$\left[-\frac{1}{2m}\hbar^2 \frac{d^2}{dx^2} + V(x)\right] \Psi(x) = E\Psi(x) \qquad \text{with}$$

(113)
$$V(x) = \begin{cases} 0 & \text{if} \quad x < a \\ V_o & \text{if} \quad -a < x < a \\ 0 & \text{if} \quad x > a \end{cases} \quad \text{where } 0 < a < \infty \text{ and } 0 < V_o < \infty$$

admits, whenever $0 < E < V_o$, a solution of the form

$$\text{(114)} \quad \Psi(x) = \begin{cases} A_- e^{ikx} + B_- e^{-ikx} & \text{if} \quad x < a \\ A e^{\kappa x} + B e^{-\kappa x} & \text{if} \quad -a < x < a \\ A_+ e^{ikx} & \text{if} \quad x > a \end{cases}$$

with

$$\text{(115)} \quad k = \left\{\frac{2mE}{\hbar^2}\right\}^{\frac{1}{2}} \quad \text{and} \quad \kappa = \left\{\frac{2m(V_o - E)}{\hbar^2}\right\}^{\frac{1}{2}}$$

where the relative ratio of the five coefficients A_-, B_-, A, B, A_+ in (114) are determined by imposing four conditions, namely that Ψ and its derivative be continuous at the boundaries $x = \pm a$. In particular, these conditions imply

$$A_- = A_+ e^{ika} \frac{1}{4ik\kappa} \left[(\kappa + ik)^2 e^{-2\kappa a} - (\kappa - ik)^2 e^{2\kappa a} \right] .$$

Then from the reverse triangle inequality $|a - b| \geq \max\{|a| - |b| , |b| - |a|\}$:

$$\left| \frac{A_-}{A_+} \right| \geq \frac{k^2 + \kappa^2}{4k\kappa} \left(e^{2\kappa a} - e^{-2\kappa a} \right) = \left[\frac{1}{2} \frac{V_o}{\sqrt{E(V_o - E)}} \right] \sinh 2\kappa a$$

Since the term [...] is independent of $\hbar$, let us emphasize the role of $\hbar$ by rewriting the above formula as:

(116) $$\frac{|A_+|^2}{|A_-|^2} \leq C \, [\sinh 2\kappa a]^{-2} \quad .$$

From the definition of $A_\pm$ in (114), the left-hand side of (116) is to be interpreted as the transmission coefficient of the barrier. In the corresponding classical model, $0 < E < V_o$ entails that this coefficient vanishes. Thus, to demand that the quantum model approximates its classical counterpart is to require that the quantum transmission coefficient be arbitrarily small, say

(117) $$\frac{|A_+|^2}{|A_-|^2} \leq C \, [\sinh 2K]^{-2}$$

with K as large as one desires. To ensure that (117) is satisfied, the computation leading to (116) shows that it is sufficient to have: $\kappa a > K$, i.e.

(118) $$\hbar < K^{-1} \, [2m(V_o - E)]^{\frac{1}{2}} \, a \quad .$$

Hence, the classical limit of the quantum system (113) now is controlled:

(i) mathematically, through the conjunction of (117) and (118);

(ii) physically, as (118) gives an estimate of its range of validity in terms of the physical quantities that characterize this system

In this sense the classical limit is similar to the non-relativistic limit: the classical description emerges from the quantum theory in the same way as Newton's mechanics emerges from Einstein's special relativity theory. The key to a proper understanding is the evaluation of the domain of validity of the approximations. Having done so, I have no qualm assuring my insurance agent that my car is not going to tunnel out of my garage, nor do I worry about relativistic red-shift when I drive my car in congested traffic. Such are the physical parameters that prevail in my car, my garage and the town where I live; compare with [Gamov, 1940] where Gamow pretends with didactic gusto that $h \simeq 1$ ergsec or $c \simeq 15$ km/hour.

2. The high temperature limit.

Following the Ehrenfests, I have repeatedly argued in this essay that in QSP, the classical regime emerges when the temperature is high enough. Typically, the quantities that tell us the regime in which we operate are similar to (βh) with $\beta = 1/kT$ (where k is the Boltzmann constant $k \simeq 1.38 \times 10^{-16}$ erg degree^{-1}).

To illustrate this point, let us review the results on the black-body radiation (subsection 2.1) and the specific heat of solids (subsection 2.3).

We saw qualitatively that if $h\nu \gg kT$ Planck's formula of 1900, here (1), reproduces (5) which had been previously established by Wien in 1896 [Wien, 1896]. Quantitatively, Paschen and Wanner [Paschen and Wanner, 1899] had verified in 1899 that Wien's formula is in agreement with laboratory data in the range of visible light, i.e. for wave length $\lambda = c/\nu$ between 4000 Å and 7000 Å, for temperatures up to 4000 K; this is what we would call today the "quantum regime". As the temperature at the surface of the sun is about 6000 K, going to much higher temperatures was then not an option. Nevertheless, as the ratio $h\nu/kT$ is concerned, raising T or lowering ν have the same effect; the latter means pushing the observation into the infrared, which was possible at the time. Indeed, the following year Lummer and Pringsheim [Lummer and Pringsheim, 1900] recorded systematic deviations from Wien's formula when the wavelength reaches the range of 12 to 18 μ (recall $1\mu = 10^{-6}$m and thus $12\mu = 12 \cdot 10^4$Å, compared with $\sim 7 \cdot 10^3$Å for visible red). This is the observation that prompted the purely classical derivation of the Rayleigh–Jones formula (6), and then Planck's interpolation between $h\nu \gg kT$ (Wien) and $h\nu \ll kT$ (Rayleigh–Jones). Experimentally, the passage from Pashen and Wanner to Lummer and Pringsheim thus marks very sharply in time (less than two years) the crossing of the boundary from the quantum to the classical regimes. These two regimes are numerically characterized by their distance $|\lambda - \lambda_{max}|$ from the wavelength λ_{max} — or equivalently the frequency ν_{max} — at which the Planck distribution (1) passes through a maximum.

As for the specific heat of solids, upon revisiting equations (8) and (9) Debye already verified that conclusion (10) can be sharpened to give the exact result

$$\text{(119)} \quad \text{where} \quad \left.\begin{array}{l} C_V = 3R\{4D(\frac{\Theta}{T}) - 3(\frac{\Theta}{T})[\exp(\frac{\Theta}{T}) - 1]^{-1}\} \\ \\ D(x) = \int_o^x dt \frac{t^3}{e^t - 1} \quad \text{and} \quad k\Theta = h\nu_o \end{array}\right\} .$$

Scholium 2 was obtained by noticing that

$$D(x) \simeq \begin{cases} 1 & \text{for} \quad 0 < x \ll 1 \\ \\ \frac{1}{5}\pi^4 x^{-3} & \text{for} \quad x \gg 1 \end{cases} .$$

To go further than this, and determine the onset of the classical regime $C_V = 3R$, requires one to notice two things. First, C_V in (119) is a monotonically increasing *universal* function of the variable Θ/T; while it cannot be written in terms of elementary functions, it can be computed numerically. Since Θ is known in term of the cut-off ν_o on the vibrational frequencies of the crystal, its value can be determined by mechanical means: for instance, at room temperature, Θ is about

100 K (for lead) and about 400 K (for aluminium), with silver and copper in between. For these, and many other metals, the measured values of the specific heat fall remarkably close to the theoretical prediction (119); cf. e.g. [Wannier, 1966, fig.13.9, p.276]. This curve shows a monotonic and smooth passage from the classical to the quantum regimes as the temperature decreases. Specifically, we can now discuss quantitatively the onset of the classical regime. The exact expression (119) entails that the first two terms in the expansion of C_V for $\Theta/T \ll 1$ give

$$(120)\quad C_V \simeq 3R\{1 - \frac{1}{20}(\frac{\Theta}{T})^2\}$$

so that at room temperature $T \simeq 300$ K, the correction to the classical value $C_V = 3R$ ranges from about 0.6 % (for lead) to about 9 % for aluminium; both of which are in good agreement with experimental data.

For other early recognitions of the emergence of the classical regime in high temperature QSP, see subsections 2.4 and 2.6 where the classical ideal gas is recovered as controlled high temperature limits of both the Bose and the Fermi quantum gases.

3. The thermodynamical limit.

As its name indicates, the thermodynamical limit is designed to elicit various macroscopic thermodynamical behaviours from microscopic mechanical models. I find it convenient to separate here the problems addressed in non-equilibrium and in equilibrium statistical physics.

a. *Non-equilibrium physics.* In the classical realm already, one appeals to the large size of the systems considered to avoid the spurious appearance of recurrences in the theoretical modeling of physical phenomena such as the thermodynamical approach to equilibrium. For instance, to buttress Boltzmann's kinetic theory of gases, the Ehrenfests proposed the so-called dog-flea model, a stochastic model later revisited by Mark Kac. This model is reviewed in [Emch and Liu, 2002, section 3.4] where the results of a computer experiment are reported, involving $N = 100$ "fleas" jumping "at random" between two "dogs": a tendency to approach equilibrium is manifest during a few hundred jumps, whereas the frequency of recurrences, which Kac showed to increase exponentially with N, is observed to occur — as regularly as to be expected — over a range of several tens of thousands of trials.

In the quantum realm, a model for an actual experiment, the nuclear free-relaxation, is solved in subsection 3.3 above. Here again the model shows an approach to equilibrium practically unaffected by a "recurrence time" that grows as 2^N, where N is the number of lattice sites in the system, a macroscopic CaF_2 crystal. Hence the empirical justification for the limit $N \to \infty$ is that the relevant time-parameter for the experimentalist is $\sim 2^{10^{23}}$, which is indeed exorbitant; accordingly, I could not discern from the laboratory [Lowe and Nordberg, 1957] any concern about putative recurrences. The supporting analytic evidence is the explicit size-correction given in equation (49) and discussed with some detail in remarks 6.

b. *Equilibrium Physics.* In equilibrium situations, the thermodynamical limit is called upon to focus on properties of matter in bulk, so to speak navigating the high seas, away from the shoals of boundary effects. This often requires some elaborate rigging, in CSP as well as in QSP.

Roughly speaking, up to the middle of the twentieth century, this was achieved by replacing sums by integrals, as in equations (13–14). As in other parts of mathematical physics, this mathematical procedure is usually well under control, albeit physics sometimes requires unusual precautions, as shown explicitly in the caveat of equation (50).

Later on, especially in the modeling of phase transitions, when the emergence of collective behaviour turned out to be essential for the understanding of the phenomena at hand, and when existence questions were raised, more sophistication was demanded. In particular, for the limit where the size of the system is allowed to go to infinity, dimension enters the play; and then, in particular, the shapes of the regions considered must be such that the ratio of the surface to the volume goes to zero: cubes are fine; sponges are not. As indicated in the various models presented in section 5, it is possible to carry out such limiting procedures explicitly and successfully. The simplest examples are lattice systems, say spins on a lattice $\mathbb{Z}^d$. Examples of continuous systems are also given in this section 5; yet, in general, such systems, say on $\mathbb{R}^d$, require extra technical care to ensure spatial uniformity and to avoid bundling effects; hence, in the latter cases the theory is not always as fully controlled as one may desire; see nevertheless [Sewell, 2002]. Interactions with extremely long range may pose further problems with regard to: (a) the definition of the limiting state; and (b) the control of the limiting time evolution; such situations were met in subsections 5.1 and 5.7.A.

4. The van Hove limit.

We encountered particular instances of this limiting procedure in subsections 3.5 and 5.4. But a more general discussion was postponed to the present section.

In a brillant transposition of a theme van Hove had heard played to justify the Born approximation in the discussion of long-time asymptotic behaviour in scattering theory, he proposed in [van Hove, 1955] a variation allowing him to characterize a regime where the time-scale of the irreversible *macroscopic* phenomena is emphasized over the time-scale of the underlying reversible, Hamiltonian *microscopic* dynamics. Van Hove's original presentations were conducted for specific models by means of perturbation techniques, carried to all orders, in which he selected for summation the "most divergent diagrams." At first, his virtuoso performances drew considerable scepticism; cf. e.g. [van Kampen, 1962]. The main problem was to isolate the conditions under which the essentials of what would become a theory may emerge from the contingent diagrammatics attached to the solution of particular models. Systematic mathematical treatments are now available to show how a joint long-time/weak-coupling limit may lead from a conservative unitary evolution to a contractive dynamical semi-group; cf. e.g. [Martin, 1979; Davies, 1976a].

In terms of the focus of this subsection, namely the control of limits that allow one to ensure that the system considered is operating in a desired regime, here exponential decay, one aims at proving a result of the following form; cf. e.g. [Martin and Emch, 1975, section 4].

There exist finite constants $\tau_o > 0$ and $C > 0$ such that for $0 \leq \lambda^2 t \leq \tau_o$:

$$(121) \quad \left| \lim_{|\Lambda|\to\infty} (\Phi, U^o_{-t} U_t \Psi)_\Lambda - (\Phi, \exp(-[\Gamma + i\Delta]\lambda^2 t)\Psi) \right| \leq \lambda C$$

where $U^o_{-t}U_t$ describes the evolution in the so-called interaction picture, with $U^o_t = \exp(-iH_o t)$, $U_t = \exp(-i[H_o + \lambda V]\, t)$; $H_o, H + \lambda V, \Delta$ are self-adjoint operators, and Γ is not only self-adjoint, but also positive so as to describe *decay* in the time range $0 \leq \tau := \lambda^2 t \leq \tau_o$. Hence the term "long-time/weak-coupling limit": when the coupling constant λ is small enough [i.e. the RHS of 121 is small] the evolution is approximated by the contractive semi-group $S(\tau) := \exp(-[\Gamma + i\Delta]\tau)$ with $\tau = \lambda^2 t \in [0, \tau_o]$, provided the time t is sufficiently large in the scale measured by $t \simeq \tau/\lambda^2$.

The separation of the total Hamiltonian $H = H_o + \lambda V$ into an "unperturbed" or "free" part H_o and an "interaction" λV must be justified. Van Hove proposed that it is to be traced back to the fact that the observables of interest in irreversible processes are macroscopic (see subsection 6.2 below), thus determining a joint spectral resolution; then H_o appears as the "diagonal" part of H in this spectral resolution. For instance, $A = \int dk A(k) a^*(k) a(k)$ and $H_o = \int dk \epsilon(k) a^*(k) a(k)$. This remark also helps justify the use of the interaction picture $U^o_{-t}U_t$ since it entails that the observables of interest are invariant under the "free" evolution. As the macroscopic observables are translation invariant, the notation $\int dk$ is used to suggest that the momentum representation corresponds to the spectral resolution in which the observables and the free Hamiltonian are diagonal.

The understanding of the van Hove limit gained in the 1970s has since been confirmed and extended; cf. e.g. [Bach *et al.*, 2000; Derezinski and Früboes, 2005] and references therein; for baselines [Davies, 1976a], [Emch and Liu, 2002, section 15.2] and [Alicki and Fannes, 2001].

Although I do not wish to elaborate on the following historical point, I may mention incidentally that the use of the interaction picture helped van Hove discern in his perturbation expansions some *characteristic features* of many-body physics by which he suggested non-equilibrium QSP differ from the QFT supporting quantum scattering theory. To this day, however, I am not sure whether van Hove's Delphian utterances have been properly digested into the corpus of contemporary mathematical physics.

I should also mention here that coupled limits have been considered also in CSP. An example is the Grad limit for classical gases in which the volume V is kept fixed, the number of molecules $N \to \infty$, and the cross-section of the molecules $\sigma := \pi\, d^2 \to 0$ (thus the volume of each molecule $v(\sim d^3) \to 0$ and the density of the gas $\rho := \frac{N}{V} \to \infty$), while the mean-free path $\lambda = \frac{V}{Nd^2}$ is kept constant; in [Grad, 1958], Grad proposed this limit as a mean to derive the Boltzmann equation. For further references relative to the latter problem, see [Uffink, 2006, section I.6.2] or [Emch and Liu, 2002, section 3.3]; and in particular, for the 2-dimensional Lorentz gas (with now $\sigma = 2d$), cf. [Martin, 1979] where it is pointed out that the Grad limit and the van Hove limit (in a form of it adapted to this model) are equivalent in one important sense: they both predict the same ratio between the macroscopic time-scale validated by observations, and the microscopic time-scale provided by the mean free time between two successive collisions.

The related philosophical issues about the roles that asymptotic reasoning plays in explanation, reduction, and emergence are cogently discussed in [Batterman, 2002a]. The above four limiting procedures may bring additional water to this epistemological mill; see already [Grad, 1967].

In closing this subsection, I should at least mention *coarse-graining*, yet another procedure that has been transferred from the classical to the quantum realm [van Kampen, 1954; Emch, 1964]. One of the reasons for not bringing it up in this essay is that I did not need it for the considerations I developed here. And the reason for this may be that I have come to believe that the primacy of coarse-graining has been largely superseded by the syntax of infinite systems which allows one to bypass several awkward issues about the relations between the micro- and macroscopic worlds; see for instance in subsection 6.3 below. Yet, as with the thermodynamical limit, coarse-graining helped explore those macroscopic properties one wishes to see emerging from finer descriptions; in so doing, it also emphasizes that distinguishing differences of *scales* or *tempi* enables smooth negotiations of such passages.

6.2 Macroscopic observables

Coming back to the general formalism, assuming that the thermodynamic limit has been taken, and concentrating on space-translations, subsection 3.5 emphasized one feature that is new to *quantum* ergodic theory. Space-averaged observables are *essential observables* in the sense of the theory of superselection sectors prompted by [Wick *et al.*, 1952], i.e. they commute with all quasi-local observables and among themselves. This is yet another classical aspect of quantum theory. The specific classical description that emerges in this manner depends on the global preparation of the system (but is insensitive to local perturbations) as the very definition — and values — of these space-averaged observables depends on the translation-invariant state φ of the system one considers. This aspect of quantum ergodic theory shows up as a direct consequence of the “locality” assumed in the Haag–Kastler axioms.

Hence it is proper to regard the emergence of a classical macroscopic description out of a quantum microscopic description as a consequence of translation-invariance and locality; cf. subsections 3.5.B and 5.7. As we saw, the passage to the thermodynamical limit and the attendant emergence of macroscopic observables allow one to discern the simultaneneous existence of several thermodynamical pure phases, as for instance the non-vanishing magnetization in zero magnetic field signals the presence of a permanent magnet. Similarly, the laboratory observation of a discontinuity in the derivative of the isotherms at the ends of the Maxwell plateau is better understood if one takes the thermodynamical limit: otherwise, the isotherms are analytic all along and the theoretical description of their experiencing so extreme a bend is simply neither convenient nor useful when considering a cup of tea. And again, nobody would claim that when receiving their drinks they recognize the ice-cubes only because the size of these is infinite ... which, mercifully, it isn't. Yet, the Landau criterion for distinguishing a solid from a fluid (see subsection 5.7.b) is strictly valid only when the thermodynamical limit is considered. This is a paradox only when the definition of limits is forgotten; here as elsewhere in physics, the key to the proper understanding of limits lies in their manifesting the emergence of qualitatively different regimes.

For time-averaged observables, the situation is more complicated. Recall some basic facts. To any time-invariant state φ the GNS construction associates a representation π_φ of the C^*-algebra $\mathcal{A}$ of quasi-local observables, and a unitary representation of the time evolution under which the von Neumann algebra obtained as the weak-operator closure of $\pi_\varphi(\mathcal{A})$, namely $\mathcal{N}_\varphi := \pi_\varphi(\mathcal{A})''$, is stable. Then, while the time-average of an observable always belongs to $\mathcal{N}_\varphi$ as does its space-average, the time-average now also belongs to the commutant $\mathcal{N}_\varphi{}' = \pi_\varphi(\mathcal{A})'$ of this algebra, and thus to its center $\mathcal{Z}_\varphi := \pi_\varphi(\mathcal{A})'' \cap \pi_\varphi(\mathcal{A})'$, *if and only if* the evolution is $\eta-$abelian. The latter condition — see equation (61) — may be satisfied in some particular models, but its status is as yet too precarious to enshrine this condition as a general "axiom" on the same footing as "locality".

In spite of the limitation just described, some of the remaining ergodic properties of observables under the time-evolution, together with some of the applications of the theory, were discussed in subsection 3.5.A.

6.3 The quantum measurement process viewed from the perspective of QSP

The technical literature on quantum measurement underwent some striking developments in the 1970s — cf. e.g. [Hepp, 1972; Bell, 1975; Whitten-Wolfe and Emch, 1976]; and also [Emch, 2003; Sewell, 2005] — in part as a consequence of the advent of the algebraic approach to QSP.

Insofar as there was a consensus on what the problem was, the original doctrine is best expounded in Wigner's careful exegesis of what he called the "orthodox" theory of von Neumann [von Neumann, 1932c]; Wigner's papers are collected in [Wigner, 1997, Part II] and [Wigner, 1995, Part II]; Wigner's positions on the subject were last stated in [Wigner, 1984]. Some of the philosophical issues are outlined in [Dickson, 2006].

A renewal in the understanding and implementation of several of the basic tenets of the doctrine was largely motivated by two critiques repeatedly advanced by Wigner himself. The first critique was that "to increase the accuracy of the measurement one has to use a very large measuring apparatus" [Wigner, 1995, p. 177] or "the large size of the apparatus appears to be essential for the possibility of a measurement" [Wigner, 1995, p. 178]. The second critique is the problem of infinite regress — the so–called Wigner's friend argument; cf. e.g. [Wigner, 1995, p. 215] — that follows from the necessity "to consider the system that has been called, so far, the apparatus, to be the object of the measurement. In other words, one will bring this apparatus into interaction with a new measuring object ... [*and so on*]"; [Wigner, 1995, pp. 208-9]. As this does not appear to be a problem with which one is usually concerned in the analysis of classical measurements, Wigner reiterated a statement he attributed to Fock, but which he said he believed to be part of the teaching of the "Copenhagen school", namely that: "Measuring instruments must be described classically"; of singular relevance to the present essay, this quote is taken from a paragraph Wigner entitled "Is the measuring apparatus macroscopic?" [Wigner, 1995, p. 205].

The reason I believe to be at the core of this awkwardness is that in Wigner's heydays, physicists were still in awe of *a perceived dichotomy* between the classical and the quantum worlds. Hence a new branch of the literature on the quantum measuring problem could develop when a solution of continuity was found that bridges these two descriptions — quantum and classical — of *the one world* in which we live. This happened when the conceptualization of the physical role of limiting proceedures came under control and, in particular, the concept of macroscopic observables was understood; see subsections 6.1 and 6.2 above, references therein, and [Landsman, 2006]. I claim that the concepts developed to deal with QSP can help construct a measuring apparatus that is described in quantum terms and yet behaves, *qua* measuring apparatus, in a classical regime. I will now indicate how at least this part of the conceptual problems associated with quantum measurement has been clarified.

Let $\mathcal{A}_S$ be the algebra of observables for the system to be measured, and let $\mathcal{B} \subset \mathcal{A}_S$ be an abelian subalgebra, the self-adjoint elements of which are the observables of interest. In the interest of formal simplicity I make here the following assumptions, parts of which are easy to dispense with.

- $\mathcal{A}_S$ contains a unit I_S and is a collection of finite-dimensional matrices.
- The spectrum of $\mathcal{B}$ is non-degenerate; hence every observable $B \in \mathcal{B}$ is of the form $B = \sum_k b_k Q_k$ with $Q_k = Q_k^*$; $Q_k Q_l = \delta_{kl} Q_k$; $\sum_k Q_k = I$; and dim $Q_k = 1$.

Initially, the system of interest is in the state $\varphi_S : A \in \mathcal{A}_S \mapsto \mathrm{Tr}\rho A \in \mathbb{C}$, and we want the measuring process to determine, for all $B \in \mathcal{B}$ the values $\varphi_S(B)$, i.e. for all k , the values $\lambda_k = \varphi_S(Q_k)$, so that we can compute $\varphi_S(B) = \sum_k b_k \lambda_k$.

For this measurement, a team of quantum engineers will be asked to build a dedicated measuring apparatus described by an algebra $\mathcal{A}_M$ with self-adjoint "pointers" M_k which are in bijective correspondence with the Q_k . They prepare this apparatus in the state φ_M . For simplicity, they assume that their $\mathcal{A}_M$ contains a unit I_M and that they arrange for $\sum_k M_k = I_M$. And finally, they try to build an interactive *Hamiltonian* mechanism such that when the system of interest and the apparatus are brought into contact the initial state $\varphi^o = \varphi_S \otimes \varphi_M$ on $\mathcal{A}_S \otimes \mathcal{A}_M$ will evolve in such a manner that the following two conditions are satisfied:

(a) concerning the measuring apparatus:

$$(122) \quad \forall M_l \ : \ \begin{cases} \varphi_M(M_l) \longrightarrow \varphi^p(M_l) = \sum_k \lambda_k \psi_k(M_l) & \text{where} \\ \\ \psi_k(M_l) = \delta_{kl} \quad \text{with no dispersion} & \end{cases} ;$$

(b) concerning the system to be measured:

$$(123) \quad \forall A_S \in \mathcal{A}_S \ : \ \begin{cases} \varphi_S(A_S) \longrightarrow \varphi^p(A_S) = \sum_k \lambda_k \varphi_k(A_S) \qquad \text{where} \\ \\ \varphi_k(A_S) = \begin{cases} \varphi_S(Q_k)^{-1} \varphi_S(Q_k A_S Q_k) & \text{when } \lambda_k \neq 0 \\ \varphi_S(A_S) & \text{when } \lambda_k = 0 \end{cases} \end{cases} .$$

Let me comment on these design requirements. Note first that (122) would deliver the values $\lambda_k = \varphi_S(Q_k)$ from which one computes the expectation values $\varphi_S(B)$ of all observables for the measurement of which the apparatus was designed. I will specify later — see (125) — what is meant by the requirement that the result of the measurement be "without dispersion", i.e. formally $\varphi^p([X - \varphi^p(X)I]^2) = 0$.

To relate the requirement (123) to the familiar textbook description of the measuring process, consider briefly the particular form it takes in the von Neumann framework where φ_S is a pure state on the algebra $\mathcal{A}_S = \mathcal{B}(\mathcal{H}_S)$ and the Q_k are one-dimensional; let $\{\Phi_k\}$ be an orthonormal basis in $\mathcal{H}_S$ with $Q_k \Phi_l = \delta_{kl} \Phi_k$; in term of this basis, one can write, without loss of generality $\varphi_S(A_S) = (\Phi_S, A_S \Phi_S)$ with $\Phi_S = \sum_k c_k \Phi_k$; and $\lambda_k = |c_k|^2$. Then (123) takes the form $\varphi^p(A_S \otimes I_M) = \mathrm{Tr}(\rho^p A_S)$ with $\rho^p = \sum_k |c_k|^2 Q_k$. Hence, viewed from $\mathcal{A}_S$, the pure state-vector Φ_S evolves to the mixed density matrix ρ^p . In this sense, (123) is the general form of the so-called von Neumann (non-selective) collapse postulate for the case where the initial state of the system is not necessarily a pure state.

Note that (122) and (123) are reduced descriptions of the evolution of the state φ^o : these requirements demand only that the evolution of special observables be followed; these special observables are: (a) the pointers M_l of the apparatus; and (b) all observables A_S pertaining to the system S . In particular, the requirement (123) would not be incompatible with a measuring process (which we denote as $\longrightarrow$) driven by a unitary evolution of the composite $\mathit{system} \cup \mathit{apparatus}$.

In line with von Neumann's "relative frequencies" view of quantum probability — explicitly inspired by von Mises [von Neumann, 1932c, fn. 156] — the general form (123) applies best to a measurement performed on a beam of particles rather than separately on individual particles. Hence — in line with the interpretation of 'states of physical systems' stated in subsection 3.1 — this description of the measuring process understands that the initial state of the system S is viewed as a summary of its preparation. For instance, in the historical Stern–Gerlach experiment, an incident *beam* of silver atoms was produced by evaporation from a heated oven; cf. [Jammer, 1966, p. 133]. Thus, what the experimentalists knew was the direction of the beam and the temperature of the oven: the latter surely a macroscopic notion! Similarly, the initial state of the measuring apparatus is viewed here as the result of its preparation; adhering to this pragmatic interpretation, one ought not to impose on the initial state of a (large!) measuring apparatus that it be pure: plainly this would require an exhorbitant amount of information to be entered in its preparation — information that ought not to be actually necessary for the adequate performance of measurements aiming to collect the simple microscopic information described by the distribution $\{\lambda_k\}$.

Due to all sorts of pesky circumstances — e.g. the recurrences present in finite systems or the intrusion of the "Wigner's friend" (introduced earlier in this subsection) — our apparatus builders would be exposed to dire frustrations, unless they be granted enough time and space so that the following idelaization is a close enough approximation — to a degree chosen in advance — of their implementation of the measuring process $\varphi^o \longrightarrow \varphi^p$, namely:

$$(124)\quad \varphi^p(X) := \lim_{t\to\infty} \lim_{|\Lambda|\to\infty} \varphi^o(\alpha_t^\Lambda[X]) \quad \text{with} \quad X = \begin{cases} A_S \otimes I_M \\ \text{or} \\ I_S \otimes M_k^\Lambda \end{cases} ,$$

choosing the pointers so that in the thermodynamical limit, $\lim_{|\Lambda|\to\infty} M_k$ exist and define 'essential' observables — in the sense of subsection 6.2 above; in particular, the reader may want to review the connection with superselection rules — i.e. observables that the orthodox theory would construe to be classical. The requirement "without dispersion" in (122) may now be specified, namely one demands that

$$(125)\quad \lim_{t\to\infty} \lim_{|\Lambda|\to\infty} \left[\varphi^o(\alpha_t[(M^k]^2)_\Lambda - \{\varphi^o(\alpha_t[(M^k])_\Lambda\}^2\right] = 0 \quad .$$

There is even an additional benefit in allowing the thermodynamical limit in (124), namely that one may demand that the experimental set-up be such that the result (124) of the measurement be empirically insensitive to local perturbations in the preparation of the initial state φ_M of the apparatus. This requirement means that $\varphi^p(X)$ in (124) do not change when the initial state φ_M of the apparatus is replaced by any state $\psi_M : A \in \mathcal{A}_M \mapsto \varphi_M(D^*AD) \in \mathbb{C}$ where D is any (quasi-)local element of $\mathcal{A}_M$ satisfying the normalization $\varphi(D^*D) = 1$; or, even more generally, by any state ψ_M normal on the von Neumann algebra $\pi_{\varphi_M}(\mathcal{A}_M)''$. Such robustness pertains to the pragmatic demand that the preparation of a large(!) measuring apparatus be reasonably simple.

Here ends — at least for the main purpose of this subsection — the list of specifications demanded from our quantum engineers when constructing a measuring device.

The contribution of algebraic QSP to the solution of the quantum measurement problem is this: the above programme can be completely implemented in the sense that specific and rigorously controllable models have been built satisfying *all* of the above specifications. These models therefore establish the applicability of the algebraic approach to the foundations of physics beyond the limitations of what Wigner called the orthodox theory. In sum, this approach encompasses the description of classical regimes unknown within the confines of the orthodox theory; cf. e.g. [Hepp, 1972; Whitten-Wolfe and Emch, 1976; Emch, 2003; Sewell, 2005] and other references listed in [Landsman, 2006, subsection 6.6].

An objection to (124), namely that real-world laboratories are finitely extended in space and in time is seductive. But it neglects the main understanding that presides over taking a limit: recall subsection 6.1 above. Here also the limit defines an asymptotic regime; thus, the control of the limiting procedures allows to take into account that good experiments do require expenses in room and allotments of time, each to be evaluated in terms of the precision to which one aims. The measuring process involves a particular instance of a general macroscopic phenomenon, the "approach to equilibrium". In subsection 6.1(3) above, I commented again on the role of the thermodynamical limit $|\Lambda| \to \infty$ in the emergence of this regime.

The role of the subsequent limit $t \to \infty$ deserves a further comment in the context of the measurement process: it does *not* say that an infinite time is required to register the result of the measurement, but rather, in accordance with our general understanding of the role of limits, the existence of the limit $t \to \infty$ asserts that for every $\epsilon > 0$, there exists a time T_ϵ that can be evaluated, and is such that the measurement has been completed *for ever, within the required precision* ϵ, when $t > T_\epsilon$. Thus in contrast with the constraint of the orthodox theory requiring that the unitary evolution be sharply interrupted at the 'end' of the measurement process, our quantum engineers do not need to make provisions for switching off the measuring device. Now, not taking first the limit $|\Lambda| \to \infty$, only requires them to review their estimate of the effects of the finite size of the apparatus; from this estimate, they evaluate how large the apparatus must be so as to allow a generous time T_Λ before which they have to switch off the measurement and avoid some nasty kickback. The *controlled* limit $|\Lambda| \to \infty$ is thus not a pragmatic limitation to the validity of the theory any more than is the theoretical implementation of the thermodynamical limit ($N \to \infty$, $|\Lambda| \to \infty$ with $D := N/|\Lambda|$ fixed) to remove astronomically long recurrences from the description of the cooloing down of your everyday cup of coffee. The description obtained in the thermodynamical limit is closer to the pragmatic account of the observed cooling down than would be its description as occuring in a finite system: the latter description would indeed be hampered by superfluous, irrelevant details. To sum up, in the actual construction of models for the measuring process, the problems that our quantum engineers encountered were not with satisfying the ancillary condition $T_\epsilon \ll \tau \ll T_\Lambda$ where τ denoted the laboratory time-scale on their wristwatch. See nevertheless [Bell, 1975].

While the models *do prove* that all the demands of the above programme are compatible, it is in the very nature of models that they *cannot* prove that

(i) the conditions of the programme are necessary to an understanding of the measuring process; nor

(ii) the conditions of the programme are sufficient, as other demands may be made, and other conditions may need to be required.

Concerning remark (i), the programme presented above emphasizes possible contributions that QSP can bring to an understanding of the quantum measurement process. One specific aim was to avoid having the theories of the measurement process beached on a conceptual sandbar between the quantum and classical worlds: the programme exploits circumstances where QSP shows how the quantum description of the one world encompasses conceptually important classical aspects. Thus the irreducible quantum/classical dichotomy has now faded into more comprehensive views, QSP being one of them. The emergence of classical behaviour in quantum theory is also one of the significant aspects of the *decoherence* programme, although the likely confluence of these two approaches has not yet gained universal acceptance. For a fair description of the latter issues, and their bearing on the measurement problem, I would recommend [Landsman, 2006]; and for a vivid and somewhat confrontational exchange on the relevance of decoherence in this context, [Anderson, 1994; Adler, 2003].

Remark (ii) above has at least two aspects. One of these aspects is that while the models that establish the internal consistency of the programme discussed in this subsection are treated with mathematical rigour, they can hardly be viewed as sufficiently realistic to satisfy our colleagues on the laboratory floor. Another aspect of the above remark (ii) on sufficiency, is that I do not know how the algebraic QSP would be helpful for formulating some of the remaining challenging questions still open in the theory of quantum measurement. If I had to single out one among these, I would direct attention first to measurements now "routinely" performed on an individual quantum system; cf. e.g. [Rauch and Werner, 2000] or [Rauch, 2005]. Whether the so-called "many worlds" and "consistent histories" approaches are really called for here is too wide a question to be addressed in this essay on QSP; cf. [Dickson, 2006; Landsman, 2006].

6.4 Mathematical physics vs. theoretical physics

Several largely unsolved problems may have been overshadowed by the abundant literature on the "return to equilibrium" of small or local deviations that are driven back to equilibrium by a thermal bath; for models of such coupled systems, see paragraphs A and C in subsection 6.4.

Most of the problems discussed below occur also in classical statistical physics; QSP offers little to alleviate them, but a little it does do, and here is how.

The first of these problems is to avoid an infinite regress: if a (small) system of interest is driven to equilibrium by a (large) thermal reservoir, whence is the reservoir getting its own canonical equilibrium and temperature? Rather than a conceptual answer to this question, the KMS condition was originally conceived as a clever, but formal, transcription — from theoretical to mathematical physics — that turned out to be a wonderfully useful organizing tool.

This very success demanded that the KMS condition be given a deeper physical justification. Substantial answers were found later, diversely expressed as several stability conditions. The latter were presented in subsection 5.4 in an order in which their formulations increasingly sound more like *bona fide* variational principles. This development is thus in line with the widely held opinion that "a variational principle is considered to be the supreme form of a law of physics" [Itô, 1987, Art. 441]. This is good, but as in other fields, a philosophical question persists as to whether any science ought to be solely, or ultimately, founded on variational principles as mechanics and so many sciences have since the eighteenth century. Theoretical physics may have offered some other considerations in this regard, such as the "big-bang" and "decoherence," but their explanatory value, consistency and adequacy remain to be proven. In the meantime, it is not unreasonable to prefer the updated variational principles with which algebraic QSP has proven able to refine their more traditional versions.

A second problem raised by the physics literature on the return to equilibrium concerns the description of global transport phenomena such as heat conduction and electric resistivity due to the interactions between electrons and phonons or random impurities in metals. Van Hove proposed a programme — of which the van Hove limiting procedure is a part, see subsection 6.1(4) above — to approach this type of question. One of the remaining problems is to produce mathematically clean arguments for the claims that are made. An even larger problem still to be fully mastered is to go beyond the contingencies of particular *ad hoc* models. This will require one to explain in physical terms amenable to a mathematical description the general microscopic properties actually responsible for a realistic delineation of the time-scales and/or regimes in which one observes such macroscopic phenomena; the first examples that come to mind are Newton's "cooling law" and Fourier's "heat" theory, i.e. the exponential temperature equilibration of temperatures and the flow of heat that governs the steady temperature distribution in materials placed between sources at different temperatures. The materials presented in this essay, particularly in subsections 3.5 [e.g. eqn. (57)] or 5.4 [e.g. eqn. (86)], exemplify some of the first steps that have been taken profitably along this road. Further, and promising but still formal, results have been obtained in [Eckman *et al.*, 1999; Bonetto *et al.*, 2000; Bach *et al.*, 2000], yet much remains to be done to bridge these with earthly concerns for an understanding that would allow one to compute realistic estimates of the value of specific material transport coefficients.

A third and perhaps more troubling problem. Time-reversal or not [Earman, 2002; Fredenhagen, 2003], even in my dreams I have not yet seen any "cosmological arrow of time" flying convincingly through the landscape of the C^*-algebraic approach developed for QSP ... but neither may such a flight be ruled out as a heretical foray into this formalism [Buchholz, 2003].

A fourth direction in which to look for extensions of the programme of QSP is concerned with situations arising far away from equilibrium.

Yet a fifth arena for investigations has opened, where a connection with the algebraic approach to QSP is emerging. It will indeed be interesting to observe whether and how the maturing mathematical theory of quantum stochastic processes [Parthasarathy, 1995; Hudson, 1998] will or may throw new light on the reduction process of statistical mechanics.

Finally, QSP has of course found most of its pragmatic confirmation in the praxis of condensed matter physics and the extension of the latter into the study of complex phenomena. However, getting enmeshed here into the technical concrete details indispensable to the full mastery of this praxis would have carried us much beyond the confines of this essay. A richly documented overview of the scope of this field of enquiry may be found in [Anderson, 1994]. Yet, as with [Feynman, 1998], such matters need to be taken up again to weave in more threads and knots as well as to incite new philosophical reflections:

Vingt fois sur le métier remettez votre ouvrage ...

BIBLIOGRAPHY

[Adler, 2003] S. Adler. Why decoherence has not solved the measurement problem: a response to P.W. Anderson. *Studies in History and Philosophy of Modern Physics,* 34:135–142, 2003. See already [Anderson, 2001].

[Aichele *et al.*, 2005] T. Aichele, U. Herzog, M. Scholz and O. Benson. *Single–photon generation and simultaneous observation of wave and particle properties,* pages 35–41 in [Khrenninkov, 2005].

[Alicki *et al.*, 1998] R. Alicki, M. Bozejko and W.A. Majewski, editors. *Quantum probability.* Banach Center Publications 43, 1998.

[Alicki and Fannes, 1994] R. Alicki and M. Fannes. Defining quantum dynamical entropy. *Letters in Mathematical Physics,* 32:75–82, 1994.

[Alicki and Fannes, 2001] R. Alicki and M. Fannes. *Quantum dynamical systems.* Oxford University Press, Oxford, New York, 2001.

[Amrein *et al.*, 1977] W.O. Amrein, J.M. Jauch and K.B. Sinha. *Scattering theory in quantum mechanics.* Benjamin, Reading MA, 1977.

[Anderson, 1994] P.W. Anderson. *A career in theoretical physics.* World Scientific, Singapore, 1994.

[Anderson, 2001] P.W. Anderson. Science: a 'dappled world' or a 'seamless web' ? *Studies in History and Philosophy of Modern Physics,* 32:487–494, 2001.

[Andries *et al.*, 1995] J. Andries, M. Fannes, P. Tuyls and R. Alicki. The dynamical entropy of the quantum Arnold cat map. *Letters in Mathematical Physics,* 35:375–383, 1995. See already [Alicki and Fannes, 1992].

[Arafune *et al.*, 2003] J. Arafune *et al.*, editors. *A garden of quanta, essays in honor of Hiroshi Ezawa.* World Scientific, Singapore, 2003.

[Araki, 1961/2] H. Araki. Einfürung in die Axiomatische Quantenfeldtheorie, unpublished lecture notes at the ETH–Zurich, 1961/2.

[Araki, 1969] H. Araki. Gibbs state of a one dimensional quantum lattice. *Communications in Mathematical Physics,* 14:120–157, 1969. See also [Araki, 1970] and [Araki, 1975].

[Araki, 1970] H. Araki. *One dimensional quantum lattice system,* pages 75-86 in [Michel and Ruelle, 1970].

[Araki, 1974] H. Araki. On the equivalence of the KMS condition and the variational principle for quantum lattice systems. *Communications in Mathematical Physics,* 38:1–10, 1974.

[Araki, 1975] H. Araki. On the uniqueness of KMS states of one–dimensional quantum lattice systems. *Communications in Mathematical Physics,* 44:1–7, 1975. See also [Kishimoto, 1976].

[Araki, 1984] H. Araki. On the XY–model on two–sided infinite chain. *Publications of the Research Institute for Mathematical Sciences, Kyoto University,* 20:277–296, 1984. See also [Araki and Barouch, 1983].

[Araki and Barouch, 1982] H. Araki and E. Barouch. On the dynamics and ergodic properties of the XY–model. *Journal of Statistical Physics,* 31:327–345, 1983. See also [Araki, 1984; Araki and Matsui, 1985; Narnhofer, 1998].

[Araki and Matsui, 1985] H. Araki and T. Matsui. Ground states of the XY–model. *Communications in Mathematical Physics,* 101:213–245, 1985.

[Araki and Miyata, 1968] H. Araki and H. Miyata. On the KMS boundary condition. *Publications of the Research Institute for Mathematical Sciences, Kyoto,* A4:373–385, 1968.

[Araki and Moriya, 2002] H. Araki and H. Moriya. Local thermodynamical stability of Fermion lattice systems. *Letters in Mathematical Physics,* 62:33–45, 2002.

[Araki and Sewell, 1977] H. Araki and G.L. Sewell. KMS conditions and local thermodynamical stability of quantum lattice systems. *Communications in Mathematical Physics,* 52:103–109, 1977.

[Araki and Woods, 1963] H. Araki and E.J. Woods. Representations of the canonical commutation relations describing a nonrelativistic infinite free Bose gas. *Journal of Mathematical Physics,* 4:637–662, 1963.

[Arnold and Avez, 1968] V.I. Arnold and A. Avez. *Ergodic problems in classical mechanics.* Benjamin, New York, 1968.

[Bach *et al.*, 2000] V. Bach, J. Fröhlich and I.M. Segal. Return to equilibrium. *Journal of Mathematical Physics,* 41:3985–4060, 2000.

[Bak, 1967] T.A. Bak, editor. *Statistical mechanics, foundations and applications; Proceedings of the IUPAP meeting, Copenhagen, 1966.* Benjamin, New York, Amsterdam, 1967.

[Bardeen *et al.*, 1957] J. Bardeen, L.N. Cooper and J.R. Schrieffer. Theory of superconductivity. *Physical Review,* 108:1175–1204, 1957.

[Barton, 1963] G. Barton. *Introduction to advanced field theory.* Wiley, New York, 1963.

[Batterman, 2002a] R.W. Batterman. *The devil is in the details: asymptotic reasoning in explanation, reduction, and emergence.* Oxford University Press, Oxford, New York, 2002.

[Batterman, 2002b] R.W. Batterman. Asymptotics and the role of minimal models. *British Journal for the Philosophy of Science,* 53:21–38, 2002.

[Bell, 1975] J.S. Bell, On wave packet reduction in the Coleman–Hepp model. *Helvetica Physica Acta,* 48:93–98, 1975.

[Beller, 1999] M. Beller. *Quantum dialogue: the making of a revolution.* University of Chicago Press, Chicago, 1999.

[Belot and Earman, 1997] G. Belot and J. Earman. Chaos out of order: quantum mechanics, the correspondence principle and chaos. *Studies in History and Philosophy of Modern Physics,* 28B:147–182, 1997.

[Benatti *et al.*, 1991a] F. Benatti, H. Narnhofer and G.L. Sewell. A non–commutative version of the Arnold cat map. *Letters in Mathematical Physics,* 21:157–172, 1991. Erratum in [Benatti *et al.*, 1991b].

[Benatti *et al.*, 1991b] F. Benatti, H. Narnhofer and G.L. Sewell. Erratum: "A non–commutative version of the Arnold cat map". *Letters in Mathematical Physics,* 22:81, 1991.

[Besicovitch, 1954] A.S. Besicovitch. *Almost periodic functions.* Dover, New York, 1954.

[Birkhoff, 1931] G.D. Birkhoff. Proof of the ergodic theorem. *Proceedings of the National Academy of Sciences (USA),* 17:656–660, 1931.

[Birkhoff, 1942] G.D. Birkhoff. What is the ergodic theorem? *American Mathematical Monthly,* 49:222–236, 1942. See also [von Neumann, 1932a].

[Bisognano and Wichmann, 1975] J.J. Bisognano and E.H. Wichmann. On the duality condition for a Hermitian scalar field. *Journal of Mathematical Physics,* 16:985–1007, 1975. See also [Bisognano and Wichmann, 1976].

[Bisognano and Wichmann, 1976] J.J. Bisognano and E.H. Wichmann. On the duality condition for quantum fields. *Journal of Mathematical Physics,* 17:303–321, 1976.

[Boltzmann, 1884] L. Boltzmann. Ableitung der Stefanschen Gesetzes betreffend die Abhängigkeit der Wärmestrahlung von Temperatur aus der electromagnetischen Lichttheorie. *Annalen der Physik und Chemie,* 22:291–294, 1884.

[Bonch-Bruevich and Tyablikov, 1962] V.L. Bonch–Bruevich and S.V. Tyablikov. *The Green function method in statistical mechanics.* North–Holland, Amsterdam, 1962.

[Bonetto *et al.*, 2000] F. Bonetto, J.L. Lebowitz and L. Rey–Bellet. *Fourier's law: a challenge to theorists,* pages 128–150 in [Fokas *et al.*, 2001].

[Borchers, 1993] H.J. Borchers. On modular inclusion and spectrum condition. *Letters in Mathematical Physics,* 27:311-323, 1993.

[Borchers, 1995] H.J. Borchers. On the use of modular groups in quantum field theory. *Annales de l'Institut Henri Poincaré,* 67:331–382, 1995.

[Borchers, 1999] H.J. Borchers. On the embedding of von Neumann subalgebra. *Communications in Mathematical Physics,* 205:69–79, 1999. See already [Borchers, 1993].

[Borchers, 2000] H.J. Borchers. On revolutionizing quantum field theory with Tomita's modular theory. *Journal of Mathematical Physics,* 41:3604–3673, 2000. See already [Borchers, 1995].

[Bose, 1924] S.N. Bose. Plancks Gesetz und Lichtquantenhypothese. *Zeitschrift für Physik,* 26:178–181, 1924.

[Brout, 1965] R.H. Brout. *Phase transitions.* Benjamin, New York, Amsterdam, 1965.

[Browder, 1976] F.E. Browder, editor. *Mathematical developments arising from Hilbert's problems,* volume 28 in *Proceedings of Symposia in Pure Mathematics,* 1976.

[Brunetti *et al.*, 2003] R. Brunetti, K. Fredenhagen and R. Verch. The generally covariant locality principle — a new paradigm for local quantum field theory. *Communications in Mathematical Physics,* 237:31–68, 2003. See also [Fredenhagen, 2003].

[Buchholz, 2003] D. Buchholz. On hot bangs and the arrow of time in relativistic quantum field theory. *Communications in Mathematical Physics,* 237:271–288, 2003.

[Buchholz and Lechner, 2004] D. Buchholz and G. Lechner. Modular nuclearity and localization. *Annales de l'Institut Henri Poincaré,* 5:1065–1080, 2004.

[Buchholz *et al.*, 2002] D. Buchholz, I. Ojima and H. Roos. Thermodynamic properties of non–equilibrium states in quantum field theory. *Annals of Physics,* 297:219–242, 2002.

[Bunge, 1967] M. Bunge, editor. *Delaware seminar in the foundations of physics.* Springer, New York, 1967.

[Caldirola, 1961] P. Caldirola, editor. *Ergodic theories.* Academic Press, New York, 1961.

[Catto *et al.*, 1998] I. Catto, C. Le Bris and P.-L. Lions, *The mathematical theory of thermodynamical limits: Thomas–Fermi type models.* Oxford University Press, Oxford, New York, 1998.

[Chandrasekhar, 1931a] S. Chandrasekhar. The maximum mass of ideal white dwarfs. *The Astrophysical Journal,* 74:81–82, 1931. See also [Chandrasekhar, 1931b], [Chandrasekhar, 1931c], [Chandrasekhar, 1932].

[Chandrasekhar, 1931b] S. Chandrasekar. The highly collapsed configuration of a stellar mass. *Monthly Notices Royal Astronomical Society,* 91:456-466, 1931.

[Chandrasekhar, 1931c] S. Chandrasekar. The density of white dwarf stars. *Philosophical Magazine,* 11:592–596, 1931.

[Chandrasekhar, 1932] S. Chandrasekar. Some remarks on the state of matter in the interior of stars. *Zeitschrift für Astrophysik,* 5:321–327, 1932.

[Chandrasekhar, 1958] S. Chandrasekhar. *An Introduction to the study of stellar structure.* Dover, New York, 1958. (1st ed. University of Chicago Press, 1939).

[Chrétien *et al.*, 1968] M. Chrétien, E.P. Gross and S. Deser, editors. *Statistical Physics, Brandeis Summer Institute, 1966.* Gordon and Breach, New York, 1968.

[Churchill and Brown, 1990] R.V. Churchill and J.W. Brown. *Complex variables and applications,* 5th ed. McGraw-Hill, New York, 1990.

[Clifton and Halvorson, 2001] R. Clifton and H. Halvorson. Are Rindler quanta real? Inequivalent particle concepts in quantum field theory. *British Journal for the Philosophy of Science,* 52:417–470, 2001.

[Cohen, 1962/1968] E.G.D. Cohen, editor. *Fundamental problems in statistical mechanics, i, ii.* North–Holland, Amsterdam, 1962, 1968.

[Connes, 2000] A. Connes. A short survey of noncommutative geometry. *Journal of Mathematical Physics,* 41:3832–3891, 2000.

[Cornfeld *et al.*, 1982] I. Cornfeld, S.V. Fomin and Ya.G. Sinai. *Ergodic theory,* Springer, Berlin, 1982.

[Davies, 1972] E.B. Davies. Diffusion for weakly coupled oscillators. *Communications in Mathematical Physics,* 27:309–325, 1972.

[Davies, 1974] E.B. Davies. Markovian master equations I. *Communications in Mathematical Physics,* 39:91–110, 1974.

[Davies, 1976a] E.B. Davies *Quantum theory of open systems,* Academic Press, London, 1976. See also [Davies, 1976b], [Davies, 1976c].

[Davies, 1976b] E.B. Davies. Markovian master equations II. *Mathematische Annalen,* 219:147–158, 1976.

[Davies, 1976c] E.B. Davies. Markovian master equations III. *Annales de l'Institut Henri Poincaré,* 11:265–273, 1976. See also [Gorini *et al.*, 1978].

[Davies, 1975] P.C.W. Davies. Scalar particle production in Schwarzschild and Rindler metric. *Journal of Physics,* A8:609–616, 1975. See also [Rindler, 1966]; [Unruh, 1976]; [Davies, 1978]; [Hawking and Page, 1983]; [Unruh and Wald, 1984]; [Wald, 1994]; [Clifton and Halverson, 2001].

[Davies, 1978] P.C.W. Davies. Thermodynamics of black holes. *Reports on Progress in Physics,* 41:1313–1355, 1978.

[de Finetti, 1937] B. de Finetti. La prévision, ses lois logiques, et ses sources objectives. *Annales de l'Institut Henri Poincaré,* 7:1–68, 1937. See also [de Finetti, 1972].

[de Finetti, 1972] B. de Finetti. *Probability, induction and statistics.* Wiley, New York, 1972.

[Debye, 1912] P. Debye. Zur Theorie der spezifischen Wärme. *Annalen der Physik,* 39:789–839, 1912.

[Derezinski and Früboes, 2005] J. Derezinski and R. Früboes. Stationary van Hove limit. *Journal of Mathematical Physics,* 46:1–15, 2005.

[Dickson, 2006] M. Dickson, This volume, chapter 4, 2006.

[Dixmier, 1957] J. Dixmier. *Les algèbres d'opérateurs dans l'espace Hilbertien (Algèbres de von Neumann).* Gauthier–Villars, Paris, 1957.

[Domb and Green, 1972] C. Domb and M.S. Green, editors. *Phase transitions and critical phenomena, I: exact results.* Academic Press, London, New York, 1972.

[Donkor *et al.*, 2003] E. Donkor, A.R. Pirich and H.E. Brandt, editors. *Quantum information and computation,* International Society for Optical Engineering (SPIE), (2003).

[Dubin and Sewell, 1970] D.A. Dubin and G.L. Sewell. Time–translations in the algebraic formulation of statistical mechanics. *Journal of Mathematical Physics,* 11:2990–2998, 1970.

[Dunford and Schwartz, 1964] N. Dunford and J.T. Schwartz. *Linear operators, part i; General theory.* Interscience, New York, 1964.

[Earman, 2002] J. Earman. What time–reversal invariance and why it matters. *International Journal for the Philosophy of Science,* 16:245–264, 2002.

[Earman and Rédei, 1996] J. Earman and M. Rédei. Why ergodic theory does not explain the success of equilibrium statistical mechanics. *British Journal for the Philosophy of Science,* 47:63–78. 1996.

[Eckman *et al.*, 1999] J.P. Eckmann, C.A. Pillet and L. Rey–Bellet. Non–equilibrium statistical mechanics of anharmonic chains coupled to two baths at different temperatures. *Communications in Mathematical Physics,* 201:657–697, 1999.

[Ehrenfest and Ehrenfest, 1911] P. Ehrenfest and T. Ehrenfest. *Begriffliche Grundlagen der statistischen Auffassungen in der Mechanik,* volume 4/22 of *Enzyklopädie der mathematischen Wissenschaften,* Teubner, Leipzig, 1911.

[Einstein, 1905a] A. Einstein. Ueber einen die Erzeugung und Verwandlung des Lichtes betreffenden heuristischen Gesichtspunkt. *Annalen der Physik,* 17:132–148, 1905; transl. in [Stachel, 1998] pp. 177–197. See also [Einstein, 1906a].

[Einstein, 1905b] A. Einstein. Ueber die von der molekularkinetischen Theorie der Wärme gefordete Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen. *Annalen der Physik,* 17:549–560, 1905; transl. in [Einstein, 1926] pp. 1–18; and in [Stachel, 1998] pp. 85–98.

[Einstein, 1906a] A. Einstein. Zur Theorie der Lichterzeugung und Lichtabsorption. *Annalen der Physik,* 20:190–206, 1906.

[Einstein, 1906b] A. Einstein. Eine neue Bestimmung der Moleküldimensionen. *Annalen der Physik,* 19:289–306, 1906; transl. in [Einstein, 1926] pp. 36–61; compare with Einstein's Ph.D. dissertation; transl. in [Stachel, 1998] pp. 45–63. For an update of the 1906 paper, see [Einstein, 1911a].

[Einstein, 1906c] A. Einstein. Zur Theorie der Brownschen Bewegung. *Annalen der Physik,* 19:371-381, 1906; transl. in [Einstein, 1926] pp. 19–35.

[Einstein, 1907] A. Einstein. Die Plancksche Theorie der Strahlung und die Theorie der spezifischen Wärme. *Annalen der Physik,* 22:180–190, 1907.

[Einstein, 1909a] A. Einstein. Zum gegenwärtigen Stand des Strahlungsproblems. *Physikalische Zeitschrift,* 10:185–193, 1909. See also [Einstein, 1909b].

[Einstein, 1909b] Ueber die Entwicklung unserer Anschauungen über das Wesen und die Konstitution der Strahlung. *Physikalische Zeitschrift,* 10:817–825, 1909.

[Einstein, 1911a] Berichtigung zu meiner Arbeit. *Annalen der Physik,* 34:591–592, 1911; transl. in [Einstein, 1926] p. 62.

[Einstein, 1911b] A. Einstein. Elementare Betrachtungen über die thermische Molekularbewegung in festen Körpern. *Annalen der Physik,* 35:679–694, 1911.

[Einstein, 1921] A. Einstein. *Geometry and experience,* (Lecture before the Prussian Academy of Sciences, January 27, 1921) pages 232–249 in [Einstein, 1954].

[Einstein, 1924] A. Einstein. Quantentheorie des einatomigen idealen Gases. *Sitzungsberichte der Deutschen Akademie der Wissenschaften, Berlin,* 1924:261–267.

[Einstein, 1926] A. Einstein. *Investigations on the theory of the Brownian movement,* R. Fürth, editor, A.D. Cowper, transl., Methuen, 1926, reprinted by Dover, New York, 1956; this is a collection of five papers published by Einstein in the years 1905–1908.

[Einstein, 1933] A. Einstein. *On the method of theoretical physics,* Herbert Spencer Lecture, 1933; in [Einstein, 1954] pp. 270–276.

[Einstein, 1954] A. Einstein. *Ideas and Opinions,* S. Bargmann, transl. and rev., Bonanza Books, New York, 1954; based *inter al.* on *Mein Weltbild,* C. Seelig, editor. Querido Verlag, Amsterdam, 1934.

[Emch, 1964] G.G. Emch. Coarse–graining in Liouville space and master equations. *Helvetica Physica Acta,* 37:532–544, 1964.

[Emch, 1972a] G.G. Emch. *Algebraic methods in statistical mechanics and quantum field theory.* Wiley, New–York, 1972.

[Emch, 1972b] G.G. Emch. *The C^*-algebraic approach to phase transitions,* pages 137–175 in [Domb and Green, 1972].

[Emch, 1976] G.G. Emch. Generalized K–Flows. *Communications in Mathematical Physics,* 49:191–215, 1976. See also [Kümmerer and Schröder, 1983].

[Emch, 1998a] G.G. Emch. *On the need to adapt de Finetti's probabilistic interpretation to QM,* pages 157–166 in [Alicki *et al.*, 1998].

[Emch, 1998b] G.G. Emch. *Chaotic dynamics in non–commutative geometry,* pages 193–205 in [Strasburger *et al.*, 1998].

[Emch, 2003] G.G. Emch. *Quantum measurement: a bridge to classical physics,* volume 5105, pages 255–264 in [Donkor *et al.*, 2003].

[Emch, 2005] G.G. Emch. Probabilistic issues in statistical mechanics. *Studies in History and Philosophy of Modern Physics,* 36:303–322, 2005. See already [Emch, 1998a].

[Emch and Guenin, 1966] G.G. Emch and M. Guenin. Gauge invariant formulation of the BCS model. *Journal of Mathematical Physics,* 7:915–921, 1966.

[Emch *et al.*, 1970] G.G. Emch, H.J.F. Knops and E.J. Verboven. Breaking of the Euclidean symmetry with an application to the theory of crystallization. *Journal of Mathematical Physics,* 11:1655–1668, 1970.

[Emch and Knops, 1970] G.G. Emch and H.J.F. Knops. Pure thermodynamical phases as extremal KMS states. *Journal of Mathematical Physics,* 11:3008–3018, 1970.

[Emch and Radin, 1971] G.G. Emch and C. Radin. Relaxation of local thermal deviation from equilibrium. *Journal of Mathematical Physics,* 12:2043–2046, 1971.

[Emch *et al.*, 1994a] G.G. Emch, H. Narnhofer, W. Thirring and G.L. Sewell. Anosov actions on non–commutative algebras. *Journal of Mathematical Physics,* 35:5582–5599, 1994.

[Emch *et al.*, 1994b] G.G. Emch, G.C. Hegerfeldt and L. Streit, editors. *On Klauder's path: a field trip.* World Scientific, Singapore, 1994.

[Emch and Liu, 2002] G.G. Emch and C. Liu. *The logic of thermostatistical physics.* Springer, Heidelberg, 2002.

[Erdélyi, 1953] A. Erdélyi. *Higher transcendental functions.* McGrawHill, New York, 1953.

[Fannes and Tuyls, 1999] M. Fannes and P. Tuyls. A continuity property of quantum dynamical entropy. *Infinite dimensional analysis, quantum probability and related topics.* 2:511–527, 1999.

[Fermi, 1926] E. Fermi. Zur Quantelung des idealen einatomigen Gases. *Zeitschrift für Physik,* 36:902-912, 1926.

[Fermi, 1927] E. Fermi. Uno modelo statistico per la determinazione d'alcune proprietà dell' atomo. *Atti della Accademia Nazionale dei Lincei, Rendiconti,* 6:602-607, 1927. See also [Fermi, 1928].

[Fermi, 1928] E. Fermi. Eine statistische Methode zur Bestimmung einiger Eigenschaften des Atomes und ihre Anwendungen auf die Theorie des periodischen Systems der Elemente. *Zeitschrift für Physik,* 48:73-79, 1928.

[Feynman, 1998] R.P. Feynman. *Statistical mechanics, a set of lectures.* Addison Wesley, Reading MA, 1998.

[Fokas *et al.*, 2001] A. Fokas, A. Grigoryan, T. Kibble and B. Zegarlinski, editors. *Mathematical Physics 2000.* International Press of Boston, Somerville MA, 2001.

[Ford *et al.*, 1965] G.W. Ford, M. Kac and P. Mazur. Statistical mechanics of assemblies of coupled oscillators. *Journal of Mathematical Physics,* 6:504–515, 1965.

[Fowler, 1926] R.H. Fowler. On dense matter. *Monthly Notices Royal Astronomical Society,* 87:114, 1926.

[Fredenhagen, 2003] K. Fredenhagen. *Locally covariant quantum theory,* in *Proceedings of the International Conference on Mathematical Physics,* Lisbon, 2003.

[Fredenhagen and Haag, 1990] K. Fredenhagen and R. Haag. On the derivation of Hawking radiation associated with the formation of a black hole. *Communications in Mathematical Physics,* 127:273–284, 1990.

[Friedrichs, 1953] K.O. Friedrichs. *Mathematical aspects of the quantum theory of fields.* Wiley, New York, 1953.

[Gamov, 1928] G. Gamow. Zur Quantentheorie des Atomkernes. *Zeitschrift für Physik,* 51:204–212, 1928.

[Gamov, 1940] G. Gamow. *Mr. Tompkins in Wonderland.* Macmillan, New York, 1940.

[Garcia–Bondia *et al.*, 2003] J.M. Garcia–Bondia, J. Varilly and H. Figueroa. *Elements of non-commutative geometry.* Birkhäuser, Boston, 2001.

[Gibbs, 1902] J.W. Gibbs. *Elementary principles of statistical mechanics, developed with special reference to the rational foundation of thermodynamics.* Scribner, New York, 1902.

[Glimm *et al.*, 1990] J. Glimm, J. Impagliazzo and I. Singer, editors. The legacy of John von Neumann. *Proceedings of Symposia in Pure Mathematics,* volume 50, 1990.

[Gorini *et al.*, 1978] V. Gorini, A. Frigerio, M. Verri, A. Kossakowski and E.C.G. Sudarshan. Properties of quantum Markovian master equations. *Reports on Mathematical Physics,* 13:149–173, 1978.

[Gouy, 1888] L. Gouy. Note sur le mouvement brownien. *Journal de Physique (2),* 7:561–564, 1888.

[Grad, 1958] H. Grad. *Principles of the kinematic theory of gases,* volume 12 of *Handbuch der Physik.* Springer, Berlin, 1958. See also [Grad, 1963].

[Grad, 1963] H. Grad. Asymptotic theory of the Boltzmann equation. *Physics of Fluids,* 6:147–181, 1963.

[Grad, 1967] H. Grad. *Levels of description in statistical mechanics and thermodynamics,* pages 49–76 in [Bunge, 1967].

[Greenleaf, 1969] F.P. Greenleaf. *Invariant mean on topological groups.* van Nostrand–Reinhold, New York, 1969.

[Gurney and Condon, 1928] R.W. Gurney and E.U. Condon. Quantum mechanics and radioactive disintegration. *Nature,* 122:439, 1928. See also [Gurney and Condon, 1929].

[Gurney and Condon, 1929] R.W. Gurney and E.U. Condon. Quantum mechanics and radioactive disintegration. *Physical Review,* 33:127–140, 129. See also [Gamov, 1928].

[Gutzwiller, 1990] M.C. Gutzwiller. *Chaos in classical and quantum mechanics.* Springer, New York, 1990.

[Haag, 1955] R. Haag. On quantum field theory. *Matematisk-Fysiske Meddelelser Danske Videnskabernes Selskab,* 29(12), 1955.

[Haag, 1959a] R. Haag. *Discussion des "axiomes" et des propriétés asymptotiques d'une théorie des champs locale avec particules composées,* in *Les problèmes mathématiques de la théorie quantique des champs,* Colloque CNRS, Lille, 1958; CNRS, Paris, 1959.

[Haag, 1959b] R. Haag. The framework of quantum field theory. *Il Nuovo Cimento,* 14:131–152, 1959.

[Haag, 1962] R. Haag. The mathematical structure of the Bardeen–Cooper–Schrieffer model. *Il Nuovo Cimento,* 25:287–299, 1962.

[Haag and Schroer, 1962] R. Haag and B. Schroer. Postulates of quantum field theory. *Journal of Mathematical Physics,* 3:248–256, 1962.

[Haag and Kastler, 1964] R. Haag and D. Kastler. An algebraic approach to quantum field theory. *Journal of Mathematical Physics,* 5:848–861, 1964.

[Haag *et al.*, 1967] R. Haag, N. Hugenholtz and M. Winnink. On the equilibrium states in quantum statistical mechanics. *Communications in Mathematical Physics,* 5:215–236, 1967.

[Haag *et al.*, 1974] R. Haag, D. Kastler and E.B. Trych–Pohlmeyer. Stability and equilibrium states. *Communications in Mathematical Physics,* 38:173–193, 1974. See [Haag and Trych–Pohlmeyer, 1977].

[Haag and Trych–Pohlmeyer, 1977] R. Haag and E.B. Trych–Pohlmeyer. Stability properties of equilibrium states. *Communications in Mathematical Physics,* 56:213–224, 1977.

[Haag *et al.*, 1994] R. Haag, H. Narnhofer and U. Stein. On quantum field theory in gravitational background. *Communications in Mathematical Physics,* 94:219–238, 1994.

[Halvorson, 2006] H. Halvorson. This volume, chapter 8, 2006.

[Hawking, 1975] S.W. Hawking. Particle creation by black holes. *Communications in Mathematical Physics,* 43:199–220, 1975.

[Hawking, 1981] R.W. Hawking. Interacting quantum fields around a black hole. *Communications in Mathematical Physics,* 80:421–442, 1981.

[Hawking and Page, 1983] S.W. Hawking and D.N. Page. Thermodynamics of black holes in anti–de Sitter space. *Communications in Mathematical Physics,* 87:577–588, 1983.

[Heisenberg, 1928] W. Heisenberg. Zur Theorie der Ferromagnetismus. *Zeitschrift für Physik,* 49:419–436, 1928.

[Hepp, 1972] K. Hepp. Quantum theory of measurement and macroscopic observables. *Helvetica Physica Acta,* 45:237–248, 1972.

[Hida, 1970] T. Hida. *Stationary stochastic processes.* Princeton University Press, Princeton NJ, 1970.

[Hilbert, 1899] D. Hilbert. *Grundlagen der Geometrie.* Teubner, Leipzig, 1899; translated by L. Unger: *Foundations of Geometry.* Open Court, La Salle IL, 1971.

[Hilbert, 1900] D. Hilbert. *Sur les problèmes futurs des mathématiques,* pages 58–114 in *Comptes rendus du deuxième congrès international des mathématiciens, Paris, 1900,* Gauthier–Villars, Paris, 1902. See also [Browder, 1976].

[Hilbert, 1918] D. Hilbert. Axiomatisches Denken. *Mathematische Annalen,* 78:405–415, 1918.

[Ho *et al.*, 1993] T.G. Ho, L.J. Landau and A.J. Wilkins. On the weak coupling limit for a Fermi gas in a random potential. *Reviews in Mathematical Physics,* 5:209–298, 1993.

[Huang, 1965] K. Huang. *Statistical mechanics.* Wiley, New York, 1965.

[Hudson, 1998] R.L. Hudson. *Quantum stochastic analysis after four decades,* pages 194–217 in *Probability towards 2000,* Springer Lecture Notes in Statistics 128, New York, 1998.

[Ingarden *et al.*, 1999] R.S. Ingarden, R. Smolukowski, S. Chandrasekhar and M. Kac. *Marian Smolukowski: his life and scientific work.* Polish Scientific Publishers PWN, Warsaw, 1999.

[Itô, 1987] K. Itô, editor. *Encyclopedic dictionary of mathematics.* The MIT Press, Cambridge MA and London England, 1987; 2nd edition of the English translation of the 3rd Japanese edition.

[Jammer, 1966] M. Jammer. *The conceptual development of quantum mechanics.* McGraw–Hill, New York, 1966.

[Jammer, 1974] M. Jammer. *The philosophy of quantum mechanics.* Wiley, New York, 1974.

[Jaynes, 1967] E.T. Jaynes. *Foundations of probability theory and statistical mechanics,* pages 77–101 in [Bunge, 1967]. See also [Jaynes, 1989].

[Jaynes, 1989] *Papers on probability, statistics and statistical mechanics.* Kluwer, Dordrecht, 1989. 1st ed. 1983.

[Jeans, 1905a] J.B. Jeans. On the partition of energy between matter and ether. *Philosophical Magazine,* 10:91–98, 1905. See also [Jeans, 1905b].

[Jeans, 1905b] J.B. Jeans. Comparison between two theories of radiation. *Nature,* 72:293, 1905.

[Josephson, 1982] B.D. Josephson. *The discovery of tunnelling supercurrents,* pages 157–164 in [Lundqvist, 1982].

[Kac, 1968] M. Kac. *Mathematical mechanisms of phase transitions,* pages in 241–305 in volume i of [Chrétien *et al.*, 1968].

[Kadison, 1990] R.V. Kadison. *Operator algebras — an overview,* pages 61–89 in [Glimm *et al.*, 1990].

[Kadison and Ringrose, 1983/1986] R.V. Kadison and J.R. Ringrose. *Fundamentals in the theory of operator algebras, i. Elementary theory; ii. Advanced theory.* Wiley, New York, 1983/1986.

[Kastler, 1976] D. Kastler. Equilibrium states of matter and operator algebras. *Istituto Nazionale di Alta Matematica, Symposia Mathematica,* 20:49–107, 1976. See also [Haag *et al.*, 1974] and [Narnhofer and Thirring, 1982].

[Kastler and Robinson, 1966] D. Kastler and D.W. Robinson. Invariant states in statistical mechanics. *Communications in Mathematical Physics,* 3:151–180, 1966. See also [Robinson and Ruelle, 1967].

[Kasler *et al.*, 1967] D. Kastler, R. Haag and L. Michel. Central decomposition of ergodic states. *Séminaire de physique théorique, Marseille,* 1967. See also [Kastler *et al.*, 1972], [Kastler, 1976], [Kastler and Robinson, 1966], [Robinson and Ruelle, 1967].

[Kastler *et al.*, 1972] D. Kastler, M. Mebkhout, G. Loupias and L. Michel. Central decomposition of invariant states with applications to the group of time translations and of Euclidean transformations in algebraic field theory. *Communications in Mathematical Physics,* 27:195–222, 1972.

[Kay, 1993] B.S. Kay. Sufficient conditions for quasifree states on spacetimes with a bifurcate Killing horizon. *Journal of Mathematical Physics,* 34:4519–4539, 1993.

[Kay and Wald, 1991] B.S. Kay and R.M. Wald. Theorems on the uniqueness and thermal properties of stationary nonsingular, quasifree states on spacetimes with a bifurcate Killing horizon. *Physics Reports,* 207:49–136, 1991. See also [Kay, 1993].

[Khinchin, 1957] A. Khinchin. *Mathematical foundations of information theory.* Dover, New York, 1957.

[Khrenninkov, 2005] A. Khrenninkov, editor. *Foundations of probability and physics.* American Institute of Physics, New York, 2005.

[Kishimoto, 1976] A. Kishimoto. On uniqueness of KMS states of one–dimensional quantum lattice systems. *Communications in Mathematical Physics,* 47:167–170, 1976.

[Kolmogorov, 1933] A.N. Kolmogorov. Grundbegriffe der Wahrscheinlichkeitsrechnung. Springer, Berlin, 1933; translated by N. Morrison: *Foundations of the theory of probability.* Chelsea, New York, 1956.

[Kossakowski *et al.*, 1977] A. Kossakowski, A. Frigerio, V. Gorini and M. Verri. Quantum detailed balance and the KMS condition. *Communications in Mathematical Physics,* 57:97–110, 1977; Erratum [Kossakowski *et al.*, 1978].

[Kossakowski *et al.*, 1978] A. Kossakowski, A. Frigerio, V. Gorini and M. Verri. Erratum: "Quantum detailed balance and the KMS condition". *Communications in Mathematical Physics,* 60:96, 1978.

[Kramers and Wannier, 1941] H.A. Kramers and G.H. Wannier. Statistics of the two–dimensional ferromagnet, i, ii. *Physical Review,* 60:252–262, 263–276, 1941.

[Kubo, 1957] R. Kubo. Statistical mechanics of irreversible processes. *Journal of the Physical Society of Japan,* 12:570–586, 1957.

[Kümmerer and Schröder, 1983] B. Kümmerer and W. Schröder. A Markov dilation of a non–quasifree Bloch evolution. *Communications in Mathematical Physics,* 90:251–262, 1983.

[Landau, 1932] L.D. Landau. On the theory of stars. *Physikalische Zeitschrift der Sowjetunion,* 1:285–288, 1932.

[Landau and Lifshitz, 1958a] L.D. Landau and E.M. Lifshitz, *Quantum mechanics.* Pergamon, London, 1958.

[Landau and Lifshitz, 1958b] L.D. Landau and E.M. Lifshitz, *Statistical physics.* Pergamon, London 1958.

[Landsman, 2006] N. Landsman. This volume, Chapter 5, 2006.

[Le Bris and Lions, 2005] C. Le Bris and P.–L. Lions. From atoms to crystals: a mathematical journey. *Bulletin of the American Mathematical Society,* 42:291–363, 2005.

[Lenard, 1970] A. Lenard, editor. *Statistical mechanics and mathematical problems,* volume 20 of *Lecture Notes in Physics.* Springer, New York, 1970.

[Lieb, 1982a] E.H. Lieb. Thomas-Fermi and related theories of atoms and molecules. *Reviews of Modern Physics,* 53:603-641, 1982; reprinted in [Lieb, 1997] pp. 259-297. See also [Lieb, 1982b].

[Lieb, 1982b] E.H. Lieb. Erratum: "Thomas-Fermi and related theories of atoms and molecules". *Reviews of Modern Physics,* 54:311, 1982; reprinted in [Lieb, 1997] p. 298.

[Lieb, 1990] E.H. Lieb. The stability of matter: from atoms to stars. *Bulletin of the American Physical Society,* 22:1-49, 1990; reprinted in [Lieb, 1997] pp. 11-59.

[Lieb, 1997] E.H. Lieb. *Selecta: The stability of matter, from atoms to stars.* Springer, Berlin, Heidelberg, 1997.

[Lieb, 2001] E.H. Lieb. *The Bose–gas: a subtle many–body problem,* pages 91-111 in [Fokas *et al.*, 2001].

[Lieb and Simon, 1977] E.H. Lieb and B. Simon. Thomas-Fermi theory of atoms, molecules and solids. *Advances in Mathematics,* 23:22-116, 1977.

[Liu and Emch, 2005] C. Liu and G.G. Emch. Explaining quantum spontaneous symmetry breaking. *Studies in History and Philosophy of Modern Physics,* 36:137–163, 2005.

[Lowe and Nordberg, 1957] I.J. Lowe and R.E. Nordberg. Free–induction decay in solids. *Physical Review,* 107:46–61, 1957.

[Lummer and Pringsheim, 1900] O. Lummer and E. Pringsheim. Ueber die Strahlung der schwarzen Körpers für lange Wellen. *Verhandlungen der Deutschen Physikalischen Gesellschaft,* 2:163–180, 1900.

[Lundqvist, 1982] S. Lundqvist, editor. *Nobel lectures 1971–1980.* World Scientific, Singapore 1982.

[Martin and Schwinger, 1959] P.C. Martin and J. Schwinger. Theory of many–particle systems. *Physical Review,* 115:1342–1373, 1959.

[Martin, 1979] Ph.A. Martin. *Modèles en mécanique des processus irréversibles,* volune 103 of *Lecture Notes in Physics.* Springer, Berlin, 1979. See also [Martin and Emch, 1975] and [Spohn, 1977].

[Martin and Emch, 1975] Ph.A. Martin and G.G. Emch. A rigourous model sustaining van Hove's phenomenon. *Helvetica Physica Acta,* 48:48–59, 1975. See also [Ho *et al.*, 1993].

[Messiah, 1960] A. Messiah. *Mécanique quantique.* Dunod, Paris, 1960.

[Michel and Ruelle, 1970] L. Michel and D. Ruelle, editors. *Systèmes à un nombre infini de degrés de liberté,* CNRS, Paris, 1970.

[Murray, 1990] F.J. Murray. *The rings of operators papers,* pages 57–60 in [Glimm *et al.*, 1990].

[Murray and von Neumann, 1936] F.J. Murray and J. von Neumann. Rings of Operators. *Annals of Mathematics,* 37:116–229, 1936.

[Murray and von Neumann, 1937] F.J. Murray and J. von Neumann. Rings of Operators ii. *Transactions of the American Mathematical Society,* 41:208–248, 1937.

[Narnhofer, 1998] H. Narnhofer. *Control of weak asymptotic abelianness with the help of the cross–product construction,* pages 331–339 in [Alicki *et al.*, 1998].

[Narnhofer, 2001] H. Narnhofer. Kolmogorov systems and Anosov systems in quantum theory. *Infinite Dimensional Analysis, Quantum Probability and Related Topics,* 4:85–119, 2001.

[Narnhofer, 2005] H. Narnhofer. Roots for quantum ergodic theory in von Neumann's work. *Reports on Mathematical Physics,* 55:93–107, 2005.

[Narnhofer and Thirring, 1982] H. Narnhofer and W. Thirring. Adiabatic theorem in quantum statistical mechanics. *Physical Review A,* 26:3646–3652, 1982. See also [Thirring, 1983a].

[Narnhofer and Thirring, 1989] H. Narnhofer and W. Thirring. Quantum K–systems. *Communications in Mathematical Physics,* 125:565–577, 1989.

[Narnhofer and Thirring, 1994a] H. Narnhofer and W. Thirring. Clustering for algebraic K–systems. *Letters in Mathematical Physics,* 30:307–316, 1994.

[Narnhofer and Thirring, 1994b] H. Narnhofer and W. Thirring. *Dynamical entropy of quantum systems and their Abelian counterpart,* pages 127–145 in [Emch *et al.*, 1994b].

[Ojima, 2003] I. Ojima. *How to formulate non–equilibrium local states in QFT – general characterization and extension to curved space–time,* pages 365–384 in [Arafune *et al.*, 2003].

[Parthasarathy, 1995] K.R. Parthasarathy. *Quantum stochastic calculus,* pages 1024–1035 in *Proceedings, International Congress of Mathematicians, Zurich, 1994,* Birkhäuser, Basel, 1995.

[Paschen and Wanner, 1899] F. Paschen and H. Wanner. Eine photometrische Methode zur Bestimmung der Exponentialconstanten der Emissionsfunktion. *Berliner Berichte,* 1899:5–11.

[Peierls, 1936] R. Peierls. On Ising's model of ferromagnetism. *Proceedings of the Cambridge Philosophical Society,* 32:477-481, 1932.

[Peierls, 1979] R. Peierls. *Surprises in theoretical physics.* Princeton University Press, Princeton NJ, 1979.

[Pitaevskii and Stringari, 2003] L. Pitaevskii and S. Stringari. *Bose–Einstein condensation.* Oxford University Press, Oxford, 2003.

[Planck, 1900a] M. Planck. Ueber eine Verbesserung der Wienschen Spektralgleichung. *Verhandlungen der Deutschen Physikalishen Gesellschaft,* 2:202-204, 1900.

[Planck, 1900b] M. Planck. Zur Theorie des Gesetzes der Energieverteilung im Normalspektrum. *Verhandlungen der Deutschen Physikalishen Gesellschaft,* 2:237–245, 1900; *ibid. Annalen der Physik,* 4:553–563, 1901.

[Planck, 1911] M. Planck. Deux récentes théories thermodynamiques: théorème de Nernst et hypothèse des quanta (Conférence prononcée le 16 décembre 1911 à la Société chimique allemande de Berlin); appended to [Planck, 1913].

[Planck, 1913] M. Planck. *Leçons de thermodynamique,* 3rd ed., Hermann, Paris 1913.

[Prigogine, 1997] I. Prigogine. *The end of certainty: time, chaos and the new laws of nature.* The Free Press, New York, 1997.

[Pusz and Woronowicz, 1978] W. Pusz and S.L. Woronowicz. Passive states and KMS states for general quantum systems. *Communications in Mathematical Physics,* 58:273–290, 1978.

[Rauch, 2005] H. Rauch. *Quantum phenomena tested by neutron interferometry,* pages 331–339 in [Khrenninkov, 2005].

[Rauch and Werner, 2000] H. Rauch and S.A. Werner. *Neutron interferometry,* Oxford Scientific Publishers, Oxford, 2000.

[Rayleigh, 1900] J.W.S. Rayleigh. Remarks upon the law of complete radiation. *Philosophical Magazine,* 49:539–540, 1900.

[Rayleigh, 1902] J.W.S. Rayleigh. The constant of radiation as calculated from molecular data. *Nature,* 72:243, 1902.

[Rindler, 1966] W. Rindler. Kruskal space and the uniformly accelerated frame. *American Journal of Physics,* 34:1174–1178, 1966.

[Robinson, 1965] D.W. Robinson. A theorem concerning positive metric. *Communications in Mathematical Physics,* 1:89–94, 1965.

[Robinson, 1971] D.W. Robinson. *The thermodynamical pressure in quantum statistical mechanics,* volume 9 of *Lecture Notes in Physics.* Springer, Berlin, 1971.

[Robinson, 1973] D.W. Robinson. Return to equilibrium. *Communications in Mathematical Physics,* 31:171–189, 1973.

[Robinson and Ruelle, 1967] D.W. Robinson and D. Ruelle. Extremal invariant states. *Annales de l'Institut Henri Poincaré,* 6:299–310, 1967. See already [Kastler and Robinson, 1966].

[Ruelle, 1968a] D. Ruelle. *Statistical mechanics.* Benjamin, New York, 1968.

[Ruelle, 1968b] D. Ruelle. Statistical mechanics of a one–dimensional lattice gas. *Communications in Mathematical Physics,* 9:267–278, 1968.

[Sakai, 1971] S. Sakai. *C^*-algebras and W^*-algebras.* Springer, New York, 1971.

[Schiff, 1955] L.I. Schiff. *Quantum mechanics,* 2nd ed., MacGraw Hill, London, 1955.

[Schrieffer, 1974] J.R. Schrieffer. *Theory of superconductivity.* Benjamin, New York, 1974.

[Schroer and Wiesbrock, 2000] B. Schroer and H.W. Wiesbrock. Modular theory and geometry. *Reviews in Mathematical Physics,* 12:301–326, 2000.

[Segal, 1990] A.F. Segal. *Paul the convert: the apostolate and apostasy of Saul the Pharisee.* Yale University Press, New Haven, 1990.

[Segal, 1947] I.E. Segal. Postulates for general quantum mechanics. *Annals of Mathematics,* 48:930–948, 1947.

[Segal, 1958] I.E. Segal. Distributions in Hilbert space and canonical systems of operators. *Transactions of the American Mathematical Society,* 88:12–41, 1958.

[Segal, 1959] I.E. Segal. Foundations of the theory of dynamical systems of infinitely many degrees of freedom, i. *Matematisk-Fysiske Meddelelser Danske Videnskabernes Selskab,* 31:2, 1959.

[Segal, 1961] I.E. Segal. Foundations of the theory of dynamical systems of infinitely many degrees of freedom, ii. *Canadian Journal of Mathematics,* 13:1–18, 1961.

[Segal, 1962] I.E. Segal. Foundations of the theory of dynamical systems of infinitely many degrees of freedom, iii. *Journal of Mathematics,* 6:500-523, 1962.

[Segal, 1963] I.E. Segal. *Mathematical problems of relativistic physics.* Lecture Notes in Applied Mathematics. American Mathematical Society, Providence RI, 1963.

[Sewell, 1973] G.L. Sewell. States and dynamics of infinitely extended physical systems. *Communications in Mathematical Physics,* 33:42–51, 1973.

[Sewell, 1977] G.L. Sewell. KMS conditions and local thermodynamical stability of quantum lattice systems, ii. *Communications in Mathematical Physics,* 55:53–61, 1977.

[Sewell, 1980a] G.L. Sewell. Relativity of temperature and the Hawking effect. *Physics Letters A,* 79:23–24, 1980. See also [Sewell, 1982a].

[Sewell, 1980b] G.L. Sewell. Stability, equilibrium and metastability in statistical mechanics. *Physics Reports,* 57:307–342, 1980.

[Sewell, 1982a] G.L. Sewell. Quantum fields on manifolds: PCT and gravitationally induced thermal states. *Annals of Physics,* 141:201–224, 1982.

[Sewell, 1982b] G.L. Sewell. W^*-dynamics of infinite quantum systems. *Letters in Mathematical Physics,* 6:209–213, 1982. See already [Dubin and Sewell, 1970] and [Sewell, 1973].

[Sewell, 2002] G.L. Sewell. *Quantum mechanics and its emergent macrophysics.* Princeton University Press, Princeton and Oxford, 2002.

[Sewell, 2005] G.L. Sewell. On the mathematical structure of quantum measurement theory. *Reports on Mathematical Physics,* 56:271–290, 2005.

[Shapiro and Teukolsky, 1983] S.L. Shapiro and S.A. Teukolsky. *Black holes, white dwarfs, and neutron stars: the physics of compact objects.* Wiley, New York, 1983.

[Smolukowski, 1906a] M. Smolukowski. Sur le chemin moyen parcouru par les molécules d'un gas et sur son rapport avec la théorie de la diffusion. *Bulletin International de l'Académie des Sciences de Cracovie,* 1906:202–213; reprinted with translation, pages 29–42 in [Ingarden *et al.*, 1999]. See also [Smolukowski, 1906b].

[Smolukowski, 1906b] M. Smolukowski. Zur kinetischen Theorie der Brownschen Molekular Bewegung und der Suspensionen. *Annalen der Physik,* 21:755–780, 1906.

[Smolukowski, 1916] M. Smolukowski. Drei Vorträge über Diffusion, Brownsche Molekularbewegung und Koagulation von Kolloidteichen. *Physikalische Zeitschrift,* 17:557–571, 585–599, 1916; reprinted with translation, pages 43–127 in [Ingarden *et al.*, 1999].

[Spohn, 1977] H. Spohn. Derivation of the transport equations for electrons moving through random impurities. *Journal of Statistical Physics,* 17:385–412, 1977.

[Stachel, 1998] J. Stachel, editor. *Einstein's miraculous year.* Princeton University Press, Princeton NJ, 1998.

[Stefan, 1879] J. Stefan. Ueber die Beziehung zwischen der Wärmestrahlung und der Temperatur. *Wiener Berichte,* 79:391–428, 1879.

[Størmer, 1969] E. Størmer. Symmetric states on infinite tensor products of C^*-algebras. *Journal of Functional Analysis,* 3:48–68, 1969.

[Strasburger *et al.*, 1998] A. Strasburger, S.T. Ali, J.-P. Antoine, J.-P. Gazeau and A. Odzijewicz, editors. *Quantization, coherent states and Poisson structures. Proceedings of the 1995 Bialowieza Workshop.* Polish Scientific Publications PWN, Warszawa, 1998.

[Streater, 1967] R.F. Streater. The Heisenberg ferromagnet as a quantum field theory. *Communications in Mathematical Physics,* 6:233–247, 1967.

[Streater and Wightman, 1964] R.F Streater and A.S. Wightman. *PCT, spin & statistics, and all that.* Benjamin, New York, 1964.

[Summers and Verch, 1996] S.J. Summers and R. Verch. Modular inclusion, the Hawking temperature and quantum field theory in curved space–time. *Letters in Mathematical Physics,* 37:145–158, 1996.

[Summers and White, 2003] S.J. Summers and R.K. White. On deriving space–time from quantum observables and states. *Communications in Mathematical Physics,* 237:203–220, 2003.

[Szasz, 1996] D. Szasz. Boltzmann's ergodic hypothesis, a conjecture for centuries. *Studia scientiarum mathematicarum Hungarica,* 31:299–322, 1996.

[Takeda, 1955] Z. Takeda. Inductive limit and infinite direct product of operator algebras. *Tohoku Mathematical Journal,* 7:68–86, 1955.

[Takesaki, 1970a] M. Takesaki. *Tomita's theory of modular algebras and its applications,* volume 128 of *Lecture Notes in Mathematics.* Springer, New York, 1970.

[Takesaki, 1970b] M. Takesaki. *States and automorphisms of operator algebras – standard representations and KMS boundary condition,* pages 205–246 in [Lenard, 1970].

[Takesaki, 1970c] M. Takesaki. Disjointness of the KMS states of different temperatures. *Communications in Mathematical Physics,* 17:33–41, 1970.

[Talman, 1968] J.C. Talman. *Special functions: a group theoretic approach; based on lectures by Eugene P. Wigner.* Benjamin, New York, 1968.

[Taylor, 1909] G.I. Taylor. Interference fringes in feeble light. *Proceedings of the Cambridge Philosophical Society, Mathematical and Physical Sciences,* 15:114–115, 1909.

[Thirring, 1968] W. Thirring. On the mathematical structure of the BCS–Model, ii. *Communications in Mathematical Physics,* 7:181–189, 1968. See already [Thirring and Wehrl, 1967].

[Thirring, 1983a] W. Thirring. On the equivalence of adiabatic invariance and KMS condition. *Foundations of Physics,* 13:409–415, 1983.

[Thirring, 1983b] W. Thirring. *Quantum mechanics of large systems.* Springer, New York, Wien, 1983.

[Thirring and Wehrl, 1967] W. Thirring and A. Wehrl. On the mathematical structure of the BCS–Model. *Communications in Mathematical Physics,* 4:303–314, 1967. See also [Thirring, 1968].

[Thomas, 1927] L.H. Thomas. The calculation of atomic fields. *Proceedings of the Cambridge Philosophical Society,* 23:542-548, 1927.

[Thorne, 1994] K.S. Thorne. *Black holes & time warps, Einstein's outrageous legacy,* Norton, New York, 1994.

[Tomita, 1967] M. Tomita. *Standard forms of von Neumann algebras.* 5th Functional Analysis Symposium, Mathematical Society of Japan, Sendai, 1967.

[Truesdell, 1945] C. Truesdell. A function which occurs in the theory of the structure of polymers. *Annals of Mathematics,* 46:144–157, 1945.

[Tuyls, 1998] P. Tuyls. *Comparing quantum dynamical entropies,* pages 411–420 in [Alicki *et al.*, 1998]. See already [Alicki and Fannes, 1994], and also [Fannes and Tuyls, 1999].

[Uffink, 1995] J. Uffink. Can the minimum entropy principle be regarded as a consistency requirement? *Studies in History and Philosophy of Modern Physics,* 26:223–261, 1995.

[Uffink, 2006] J. Uffink. This volume, Chapter 9, 2006.

[Uhlenbeck, 1967] G.E. Uhlenbeck. *Discusssion comment,* page 56 in [Bak, 1967].

[Uhlenbeck, 1968] G.E. Uhlenbeck. *An outline of statistical mechanics,* pages 1–29 in volume ii of [Cohen, 1962/1968].

[Unruh, 1976] W.G. Unruh. Notes on black–hole evaporation. *Physical Review D,* 14:870–892, 1976.

[Unruh and Wald, 1984] W.G. Unruh and R.M. Wald. What happens when an accelarating observer detects a Rindler particle? *Physical Review D,* 29:1047–1056, 1984.

[van Hove, 1952] L. van Hove. Les difficultés de divergences sur un modèle particulier de champ quantifié. *Physica,* 18:145–159, 1952.

[van Hove, 1955] L. van Hove. Quantum mechanical perturbations giving rise to a statistical transport equation. *Physica,* 21:517–540, 1955.

[van Hove, 1957] L. van Hove. The approach to equilibrium in quantum statistics. *Physica,* 23:441–480, 1957.

[van Hove, 1959] L. van Hove. The ergodic behaviour of quantum many–body systems. *Physica,* 25:268–276, 1959.

[van Hove, 1961] L. van Hove. *The problem of master equations,* pages 155–169 in [Caldirola, 1961].

[van Hove, 1962] L. van Hove. *Master equation and the approach to equilibrium for quantum systems,* pages 157–172 in volume i of [Cohen, 1962/1968].

[van Kampen, 1954] N.G. van Kampen. Quantum statistics of irreversible processes. *Physica,* 20:603–622, 1954. See also [van Kampen, 1988].

[van Kampen, 1962] N.G. van Kampen. *Fundamental problems in statistical mechanics of irreversible processes,* pages 173–202 in volume i of [Cohen, 1962/1968].

[van Kampen, 1967] N.G. van Kampen. *Discussion remark,* page 67 in [Bak, 1967].

[van Kampen, 1988] N.G. van Kampen. Ten theorems about quantum mechanical measurements. *Physica A,* 153:97–113, 1988.

[Verboven, 1967] E.J. Verboven. *Quantum thermodynamics of an infinite system of harmonic oscillators,* pages 49–54 in [Bak, 1967].

[Vilenkin, 1968] N. Vilenkin. *Special functions and the theory of group representations,* volume 22 of American Mathematical Society Translations, Providence RI, 1968.

[von Mises, 1928] R. von Mises. *Wahrscheinlichkeit, Statistik und ihre Wahrheit,* Springer, Wien, 1928. See also [von Mises, 1957].

[von Mises, 1957] R. von Mises. *Probability, statistics and truth,* McMillan, New York, 1957; prepared by H. Geringer from the 3rd German ed. (1951).

[von Neumann, 1932a] J. von Neumann. Proof of the quasi–ergodic theorem. *Proceedings of the National Academy of Sciences (USA),* 18:70–83, 1932. See also [von Neumann, 1932b].

[von Neumann, 1932b] J. von Neumann. Physical applications of the ergodic hypothesis. *Proceedings of the National Academy of Sciences (USA),* 18:263–266, 1932. See also [Birkhoff, 1931].

[von Neumann, 1932c] J. von Neumann. *Mathematische Grundlagen der Quantenmechanik.* Springer, Berlin, 1932; translated by R.T. Beyer: *Mathematical foundations of quantum mechanics,* Princeton University Press, Princeton NJ, 1955; page references are to this edition.

[Wald, 1977] R.M. Wald. *Space, time, and gravity: the theory of big bang and black holes.* University of Chicago Press, Chicago, 1977.

[Wald, 1994] R.M. Wald. *Quantum field theory in curved spacetime and black hole thermodynamics.* University of Chicago Press, Chicago, 1994. For an elementary introduction, see [Wald, 1977].

[Wannier, 1966] G.H. Wannier. *Statistical physics.* John Wiley, New York, 1966.

[Weinberg, 1972] S. Weinberg. *Gravitation and cosmology, principles and applications of the general theory of relativity.* Wiley, New York, 1972.

[Weiss, 1907] P. Weiss. L'hypothèse du champ moléculaire et la propriété ferromagnétique. *Journal de Physique,* 6:661–690, 1907.

[Whittaker and Watson, 1927] A.N. Whittaker and G.N. Watson, *A course in modern analysis,* 4th ed., Cambridge University Press, London, 1927.

[Whitten-Wolfe and Emch, 1976] B. Whitten-Wolfe and G.G. Emch. A mechanical quantum measuring process. *Helvetica Physica Acta,* 49:45–55, 1976.

[Wick *et al.*, 1952] G.C. Wick, A.S. Wightman and E.P. Wigner. Intrinsic parity of elementary particles. *Physical Review,* 88:101–105, 1952.

[Wien, 1894] W. Wien. Temperatur und Entropie der Strahlung. *Annalen der Physik,* 52:132–165, 1894.

[Wien, 1896] W. Wien. Ueber die Energieverteilung im Emissionspektrum eines schwarzen Körpers. *Annalen der Physik,* 58:662–669, 1896.

[Wiesbrock, 1997] H.W. Wiesbrock. Half–sided modular inclusions of von Neumann algebras. *Communications in Mathematical Physics,* 184:583–685, 1997. See also [Schroer and Wiesbrock, 2000].

[Wightman and Schweber, 1955] A.S. Wightman and S. Schweber. Configuration space methods in relativistic quantum field theory. *Physical Review,* 98:812–837, 1955.

[Wigner, 1931] E.P. Wigner. *Gruppentheorie und ihre Anwendung auf die Quantenmechanik der Atomspektren.* Vieweg, Braunschweig, 1931; translated by J.C. Griffin: *Group theory and its application to the quantum mechanics of atomic spectra.* Academic Press, New York, 1959.

[Wigner, 1962] E.P. Wigner. *Oral admonition to young researchers.* Private communication (1962), which I understand many others have had to face, before and since.

[Wigner, 1984] E.P. Wigner. *Review of the quantum–mechanical measurement problem,* pages 63–82 in *Science, computers and the information onslaught,* D.M. Kerr *et al.*, editors, Academic Press, New York, 1984; reprinted in [Wigner, 1995] pp. 225–244.

[Wigner, 1995] E.P. Wigner. *Philosophical reflections and syntheses,* volume vi of *The collected works of Eugene Paul Wigner.* Springer, Heidelberg, 1995 (soft-cover, 1997). Annotated by G.G. Emch.

[Wigner, 1997] E.P. Wigner. *Foundations of quantum mechanics,* part ii of volume iii of *The collected works of Eugene Paul Wigner.* Springer, Heidelberg, 1997. Annotated by A. Shimony.

[Winnink, 1967] M. Winnink. *On the equilibrium states of infinite quantum systems at* $T \neq 0$, pages 59–66 in [Bak, 1967].

[Zeh, 1970] H.D. Zeh. On the interpretation of measurement in quantum theory. *Foundations of Physics,* 1:69–76, 1970.

ISSUES IN THE PHILOSOPHY OF COSMOLOGY

George F R Ellis

1 INTRODUCTION

Cosmology is the study of the large-scale structure of the Universe, where '*the Universe*' means all that exists in a physical sense [Harrison, 2000]. This is to be distinguished from *the Observable Universe*, namely that part of the Universe containing matter accessible to our astronomical observations, which is a subset of the Universe proper. Thus cosmology considers the vast domain of galaxies, clusters of galaxies, quasi-stellar objects, etc., observable in the sky by use of telescopes of all kinds, examining their nature, distribution, origins, and relation to their larger environment. *Observational cosmology* [Hoyle, 1960; Kristian and Sachs, 1966; Gunn *et al.*, 1978; Sandage *et al.*, 1993; Bothun, 1998] aims to determine the large-scale geometry of the observable universe and the distribution of matter in it from observations of radiation emitted by distant objects, while *physical cosmology* [Peebles, 1971; Sciama, 1971; Weinberg, 1972; Silk, 2001; Perkins, 2005; Dodelson, 2003] is the study of interactions during the expansion of the universe in its early hot big bang phase, and *astrophysical cosmology* [Sciama, 1971; Peebles, 1993b; Padmanbhan, 1993; Rees, 1995; Dodelson, 2003] studies the resulting later development of large-scale structures such as galaxies and clusters of galaxies. Various forms of quantum cosmology (see e.g. [Hawking, 1993; Gibbons *et al.*, 2003; Copeland *et al.*, 2005]) and studies of particle physics aspects of cosmology [Kolb and Turner, 1990; Peacock, 1999; Allday, 2002; Perkins, 2005; Dodelson, 2003] attempt to characterize the epochs before the hot big bang phase. These studies function in a mainly symbiotic way, each informing and supplementing the others to create an overall cosmological theory of the origin and evolution of the physical universe [Bondi, 1960; Harrison, 2000; Silk, 1997].

A unique role of the universe is in creating the environment in which galaxies, stars, and planets develop, thus providing a setting in which local physics and chemistry can function in a way that enables the evolution of life on planets such as the Earth. If the cosmological environment were substantially different, local conditions would be different and in most cases we would not be here [Carr and Rees, 1979; Davies, 1982; Barrow and Tipler, 1984; Tegmark, 1998; Rees, 1999] — indeed no biological evolution at all would have taken place. Thus cosmology is of substantial interest to the whole of the scientific endeavor, for it sets the

Handbook of the Philosophy of Science. Philosophy of Physics
Volume editors: Jeremy Butterfield and John Earman
General editors: Dov M. Gabbay, Paul Thagard and John Woods
© 2007 Elsevier B.V. All rights reserved.

framework for the rest of science, and indeed for the very existence of observers and scientists. It is unique as the ultimate historical/geographical science.

Cosmology as a serious scientific study began with the discovery of Einstein's static universe in 1917, followed by the key observational discovery of the linear redshift-distance relation by Hubble in 1929, indicating the expansion of the universe, and the development of theoretical understanding of the geometry and dynamics of the non-static Friedmann-Lemaître models with their Robertson-Walker geometry [North, 1965; Berendzen *et al.*, 1976; Smith, 1982; Ellis, 1989; Kragh, 1996]. It has been transformed in the past decades into a mainstream branch of physics [Barnett *et al.*, 1996; Nilsson *et al.*, 1991] by the linking of nuclear and particle physics theories to observable features of the cosmos [Weinberg, 1972; Kolb and Turner, 1990; Peacock, 1999; Allday, 2002; Dodelson, 2003], and into an important part of astronomy because of the massive flow of new astronomical data becoming available [Gunn *et al.*, 1978; Harwit, 1984; Bothun, 1998], particularly through new ground-based telescopes such as Keck and through balloon and satellite observatories such as the Hubble Space telescope (optical and ultraviolet), IRAS (infra-red), ROSAT (x-ray), and COBE and WMAP (microwave). Thus the subject has progressed from a mainly mathematical and even philosophical exercise to an important part of mainstream science, with a well-established standard model confirmed by various strands of evidence [Weinberg, 1972; Peebles *et al.*, 1991; Silk, 1997; Peacock, 1999; Dodelson, 2003]. Nevertheless because of its nature, it is different from any other branch of the natural sciences, its unique features playing themselves out in the ongoing interaction between speculation, theory, and observation.

Cosmology's major difference from the other sciences is the uniqueness of its object of study — the Universe as a whole [McCrea, 1953; McCrea, 1960; Munitz, 1962] — together with its role as the background for all the rest of physics and science, the resulting problems being accentuated by the vast scale of the universe and by the extreme energies occurring in the very early universe. We are unable to manipulate in any way its originating conditions, and there are limitations on our ability to observe both to very distant regions and to very early times. Additionally, there are limits to our ability to test the physics relevant at the earliest epochs. Consequently it is inevitable that (as is also the case for the other historical sciences) philosophical choices will to some degree shape the nature of cosmological theory, particularly when it moves beyond the purely descriptive to an explanatory role [Matravers *et al.*, 1995] — that move being central to its impressive progress in recent decades. These philosophical choices will strongly influence the resulting understanding, and even more so if we pursue a theory with more ambitious explanatory aims.

After a substantial outline of present day cosmology in Section 2, these issues will be explored in the subsequent sections, based on a series of thirty-four *Theses* clustered around nine key aspects of the nature of cosmology, broadly speaking relating to geometry, physics, and philosophy, that frame the context of the philosophical issues facing cosmology and its relation to local physics. I believe this formulation helps focus on specific issues of importance in this relation. To those who believe cosmology is simply about determining a number of physical parameters, this will seem a vastly over-complicated approach; but then a major purpose of this paper is precisely to counter such simplistic visions of the nature of cosmology. For other reports on the philosophy of cosmology, see [McCrea, 1970; Munitz, 1962; Ellis, 1991; Leslie, 1994; Leslie, 1998].

2 OUTLINE OF COSMOLOGY

A series of basic features of present day cosmology are now well established. Decades of painstaking work has established the distances of galaxies and hence the huge scale of the universe, as well as the basic feature that the universe is expanding and evolving; the old dream of a static universe is unviable [Ellis, 1990]. Cosmology proceeds by assuming *the laws of physics are the same everywhere, and underlie the evolution of the universe*. The dominant role of gravity, despite its weakness, then arises from the fact that it is the only known force acting effectively on astronomical scales (the other known long-range force is electromagnetism, but in this case negative charges balance out positive charges, leaving no resultant large-scale effect). Consequently, cosmological theory describing all but the very earliest times is based on the classical relativistic theory of gravitation, namely Einstein's General Theory of Relativity [Malament, 2006], with the matter present determining space-time curvature and hence the evolution of the universe. The way this works out in any particular situation depends on the nature of the matter/fields present, described by their effective equations of state and interaction potentials.

The survey of cosmology in this section looks successively at the basic models of cosmology; the hot big bang; cosmological observations, including the Cosmic Background Radiation anisotropy spectrum; causal and visual horizons, and their implications; recent theoretical developments (including inflation); the very early universe; and the present concordance model, which includes both dark matter and dark energy.

2.1 Basic Theory

Cosmology starts by assuming that *the large-scale evolution of spacetime can be determined by applying Einstein's field equations of Gravitation* ('EFE') *everywhere*: global evolution will follow from local physics. The standard models of cosmology [Robertson, 1933; Ehlers, 1993; Weinberg, 1972; Hawking and Ellis, 1973] are based on the assumption that once one has averaged over a large enough physical scale, *isotropy is observed by all fundamental observers* (the preferred family of observers associated with the average motion of matter in the universe). When this isotropy is exact, *the universe is spatially homogeneous as well as isotropic* [Walker, 1944; Ehlers, 1993; Ellis, 1971a]. The matter motion is then along irrotational and shear-free geodesic curves with tangent vector u^a, implying the existence of a canonical time-variable t obeying $u_a = -t_{,a}$. The Robertson-Walker ('RW') geometries used to describe the large-scale structure of the universe [Robertson, 1935; Walker, 1936] embody these symmetries exactly. Consequently they are conformally flat, that is, the Weyl tensor is zero:

$$(1) \quad C_{ijkl} := R_{ijkl} + \tfrac{1}{2}(R_{ik}g_{jl} + R_{jl}g_{ik} - R_{il}g_{jk} - R_{jk}g_{il}) - \tfrac{1}{6}R(g_{ik}g_{jl} - g_{il}g_{jk}) = 0;$$

this tensor represents the free gravitational field, enabling non-local effects such as tidal forces and gravitational waves which do not occur in the exact RW geometries.

Comoving coordinates can be chosen so that the metric takes the form:

$$(2) \quad ds^2 = -dt^2 + S^2(t)\, d\sigma^2, \quad u^a = \delta^a{}_0 \ (a = 0, 1, 2, 3)$$

where $S(t)$ is the time-dependent scale factor, and the worldlines with tangent vector $u^a = dx^a/dt$ represent the histories of fundamental observers. The space sections $\{t = const\}$ are surfaces of homogeneity and have maximal symmetry: they are 3-spaces of constant curvature $K = k/S^2(t)$ where k is the sign of K. The normalized metric $d\sigma^2$ characterizes a 3-space of normalized constant curvature k; coordinates (r, θ, ϕ) can be chosen such that

$$(3) \quad d\sigma^2 = dr^2 + f^2(r)\left(d\theta^2 + \sin^2\theta d\phi^2\right)$$

where $f(r) = \{\sin r,\ r,\ \sinh r\}$ if $k = \{+1,\ 0,\ -1\}$ respectively. The rate of expansion at any time t is characterised by the *Hubble parameter* $H(t) = \dot{S}/S$.

To determine the metric's evolution in time, one applies the Einstein Field Equations ('EFE'), showing the effect of matter on space-time curvature, to the metric (2,3). Because of local isotropy, the matter tensor T_{ab} necessarily takes a perfect fluid form relative to the preferred worldlines with tangent vector u^a:

$$(4) \quad T_{ab} = (\mu + p/c^2)u_a u_b + (p/c^2)g_{ab}$$

(c is the speed of light). The energy density $\mu(t)$ and pressure term $p(t)/c^2$ are the timelike and spacelike eigenvalues of T_{ab}. The integrability conditions for the EFE are the *energy-density conservation equation*

$$(5) \quad T^{ab}{}_{;b} = 0 \Leftrightarrow \dot{\mu} + (\mu + p/c^2)3\dot{S}/S = 0\,.$$

This becomes determinate when a suitable equation of state function $w := pc^2/\mu$ relates the pressure p to the energy density μ and temperature T: $p = w(\mu, T)\mu/c^2$ (w may or may not be constant). Baryons have $\{p_{bar} = 0 \Leftrightarrow w = 0\}$ and radiation has $\{p_{rad}c^2 = \mu_{rad}/3 \Leftrightarrow w = 1/3,\ \mu_{rad} = aT^4_{rad}\}$, which by (5) imply

$$(6) \quad \mu_{bar} \propto S^{-3},\ \mu_{rad} \propto S^{-4},\ T_{rad} \propto S^{-1}.$$

The scale factor $S(t)$ obeys the *Raychaudhuri equation*

$$(7) \quad 3\ddot{S}/S = -\frac{1}{2}\kappa(\mu + 3p/c^2) + \Lambda,$$

where κ is the gravitational constant and Λ the cosmological constant.[1] This shows that the active gravitational mass density of the matter and fields present is $\mu_{grav} := \mu + 3p/c^2$. For ordinary matter this will be positive:

$$(8) \quad \mu + 3p/c^2 > 0 \;\Leftrightarrow\; w > -1/3$$

(the 'Strong Energy Condition'), so ordinary matter will tend to cause the universe to decelerate ($\ddot{S} < 0$). It is also apparent that a positive cosmological constant on its own will cause an accelerating expansion ($\ddot{S} > 0$). When matter and a cosmological constant are both present, either result may occur depending on which effect is dominant. The first integral of equations (5, 7) when $\dot{S} \neq 0$ is the *Friedmann equation*

$$(9) \quad \frac{\dot{S}^2}{S^2} = \frac{\kappa\mu}{3} + \frac{\Lambda}{3} - \frac{k}{S^2}.$$

This is just the Gauss equation relating the 3-space curvature to the 4-space curvature, showing how matter directly causes a curvature of 3-spaces [Ehlers, 1993; Ellis, 1971a]. Because of the spacetime symmetries, the ten EFE are equivalent to the two equations (7, 9). Models of this kind, that is with a Robertson-Walker ('RW') geometry with metric (2, 3) and dynamics governed by equations (5, 7, 9), are called *Friedmann-Lemaître universes* ('FL' for short). The Friedmann equation (9) controls the expansion of the universe, and the conservation equation (5) controls the density of matter as the universe expands; when $\dot{S} \neq 0$, equation (7) will necessarily hold if (5, 9) are both satisfied.

Given a determinate matter description (specifying the equation of state $w = w(\mu, T)$ explicitly or implicitly) for each matter component, the existence and uniqueness of solutions follows both for a single matter component and for a combination of different kinds of matter, for example $\mu = \mu_{bar} + \mu_{rad} + \mu_{cdm} + \mu_{\nu}$ where we include cold dark matter (cdm) and neutrinos (ν). Initial data for such solutions at an arbitrary time t_0 (eg. today) consists of,

- The *Hubble constant* $H_0 := (\dot{S}/S)_0 = 100h$ km/sec/Mpc;
- A dimensionless *density parameter* $\Omega_{i0} := \kappa\mu_{i0}/3H_0^2$ for each type of matter present (labelled by i);
- If $\Lambda \neq 0$, either $\Omega_{\Lambda 0} := \Lambda/3H_0^2$, or the dimensionless *deceleration parameter* $q_0 := -(\ddot{S}/S)_0 H_0^{-2}$.

Given the equations of state for the matter, this data then determines a unique solution $\{S(t), \mu(t)\}$, i.e. a unique corresponding universe history. The total matter density is the sum of the terms Ω_{i0} for each type of matter present, for example

$$(10) \quad \Omega_{m0} = \Omega_{bar0} + \Omega_{rad0} + \Omega_{cdm0} + \Omega_{\nu 0},$$

[1] A cosmological constant can also be regarded as a fluid with pressure p related to the energy density μ by $\{p = -\mu c^2 \Leftrightarrow w = -1\}$. For the history of the cosmological constant, see [Earman, 2001; Earman, 2003].

and the total density parameter Ω_0 is the sum of that for matter and for the cosmological constant:

(11) $\Omega_0 = \Omega_{m0} + \Omega_{\Lambda 0}$.

Evaluating the Raychaudhuri equation (7) at the present time gives an important relation between these parameters: when the pressure term p/c^2 can be ignored relative to the matter term μ (as is plausible at the present time),[2]

(12) $q_0 = \frac{1}{2}\,\Omega_{m0} - \Omega_{\Lambda 0}$.

This shows that a cosmological constant Λ can cause an acceleration (negative q_0); if it vanishes, the expression simplifies: $\Lambda = 0 \Rightarrow q = \frac{1}{2}\,\Omega_{m0}$, showing how matter causes a deceleration of the universe. Evaluating the Friedmann equation (9) at the time t_0, the spatial curvature is

(13) $K_0 := k/S_0^2 = H_0^2\,(\Omega_0 - 1)$.

The value $\Omega_0 = 1$ corresponds to spatially flat universes ($K_0 = 0$), separating models with positive spatial curvature ($\Omega_0 > 1 \Leftrightarrow K_0 > 0$) from those with negative spatial curvature ($\Omega_0 < 1 \Leftrightarrow K_0 < 0$).

The FL models are the standard models of modern cosmology, surprisingly effective in view of their extreme geometrical simplicity. One of their great strengths is their explanatory role in terms of making explicit the way the local gravitational effect of matter and radiation determines the evolution of the universe as a whole, this in turn forming the dynamic background for local physics (including the evolution of the matter and radiation).

2.1.1 The basic solutions

For baryons (pressure-free matter) and non-interacting radiation, the Friedmann equation (9) takes the form

(14) $$\frac{3\dot{S}^2}{S^2} = \frac{A}{S^3} + \frac{B}{S^4} + \frac{\Lambda}{3} - \frac{3k}{S^2}$$

where $A := \kappa\mu_{bar0}S_0^3$ and $B := \kappa\mu_{rad0}S_0^4$. The behaviour depends on the cosmological constant Λ [Robertson, 1933; Rindler, 2001].

When $\Lambda = 0$, the universe starts off at a very dense initial state — according to the classical theory, an initial singularity where the density and curvature go infinite (see Sec. 2.1.2). Its future fate depends on the value of the spatial curvature, or equivalently the density parameter Ω_0. The universe expands forever if $\{k = 0 \Leftrightarrow \Omega_0 = 1\}$ or $\{k < 0 \Leftrightarrow \Omega_0 < 1\}$, but collapses to a future singularity if $\{k > 0 \Leftrightarrow \Omega_0 > 1\}$. Thus $\Omega_0 = 1$ *corresponds to the critical density* μ_{crit} *separating* $\Lambda = 0$ *FL models that recollapse in the future from those that expand forever*, and Ω_0 is just the ratio of the matter density to this critical density:

[2]Assuming we represent 'dark energy' (Sec. 2.3.6) as a cosmological constant.

(15) $\{\Omega_{crit} = 1 \Leftrightarrow \kappa\mu_{crit} = 3H_0^2\} \;\Rightarrow\; \Omega_0 := \kappa\mu_0/3H_0^2 = \mu_0/\mu_{crit}$.

When $\Lambda < 0$, all solutions start at a singularity and recollapse.

When $\Lambda > 0$, if $k = 0$ or $k = -1$ all solutions start at a singularity and expand forever. If $k = +1$ there can again be models with a singular start, either expanding forever or collapsing to a future singularity. However in this case a static solution (the Einstein static universe) is also possible, as well as models asymptotic to this static state in either the future or the past. Furthermore models with $k = +1$ can bounce (collapsing from infinity to a minimum radius and re-expanding).

The dynamical behaviour of these models has been investigated in depth: first for dust plus a cosmological constant [Robertson, 1933; Rindler, 2001], followed by perfect fluids, fluids with bulk viscosity, kinetic theory solutions, and scalar field solutions. Current models employ a realistic mixture of matter components (baryons, radiation, neutrinos, cold dark matter, a scalar field, and perhaps a cosmological constant). Informative phase planes show clearly the way higher symmetry (self-similar) models act as attractors and saddle points for the other models [Madsen and Ellis, 1988; Ehlers and Rindler, 1989].

The simplest expanding solutions are the following:

1. The *Einstein-de Sitter* model, for which $\{p = 0,\ \Lambda = 0,\ k = 0\} \Rightarrow \Omega_0 = 1$. This is the simplest expanding non-empty solution:

   (16) $S(t) = C\,t^{2/3}$

   starting from a singular state at time $t = 0$ (C is an arbitrary constant). Its age (the proper time since the start of the universe) when the Hubble constant takes the value H_0 is $\tau_0 = \frac{2}{3H_0}$. This is a good model of the expansion of the universe since radiation domination ended until the recent times when a cosmological constant started to dominate the expansion. It is also a good model of the far future universe if $k = 0$ and $\Lambda = 0$.

2. The *Milne* model, for which $\{\mu = p = 0,\ \Lambda = 0,\ k = -1\} \Rightarrow \Omega_0 = 0$, giving a linearly expanding empty solution:

   (17) $S(t) = C\,t.$

   This is just flat space-time as seen by a uniformly expanding set of observers [Rindler, 2001, pp. 360-363], singular at $t = 0$. Its age is $\tau_0 = \frac{1}{H_0}$. It is a good model of the far future universe if $k < 0$ and $\Lambda = 0$.

3. The *de Sitter* universe, for which $\{\mu = p = 0,\ \Lambda \neq 0,\ k = 0\} \Rightarrow \Omega_0 = 0$, giving the steady state expanding empty solution:[3]

   (18) $S(t) = C\,\exp(Ht)$,

[3]The Steady State universe of Bondi, Hold and Hoyle [Bondi, 1960] utilised this metric, but was non-empty as they abandoned the EFE.

where C and H are constants. As the expansion rate is constant forever, there is no start and its age is infinite.[4] It is a good model of the far future universe for those cases which expand forever with $\Lambda > 0$. It can alternatively be understood as a solution with $\Lambda = 0$ and containing matter with the exceptional equation of state $\mu + p/c^2 = 0$. There are other RW forms of the de Sitter Universe: a geodesically complete form with $k = +1$, $S(t) = S_0 \cosh Ht$ (a regular bounce), and another geodesically incomplete form with $k = -1$, $S(t) = S_0 \sinh Ht$ (a singular start). This lack of uniqueness is possible because *this is a spacetime of constant curvature*, with no preferred timelike directions or space sections [Schrödinger, 1956; Hawking and Ellis, 1973; Rindler, 2001].[5]

2.1.2 An initial singularity?

The above are specific models: what can one say generically? When the inequality (8) is satisfied, one obtains directly from the Raychaudhuri equation (7) the

> **Friedmann-Lemaître Universe Singularity Theorem** [Ehlers, 1993; Ellis, 1971a]: In a FL universe with $\Lambda \leq 0$ and $\mu + 3p/c^2 > 0$ at all times, at any instant t_0 when $H_0 \equiv (\dot{S}/S)_0 > 0$ there is a finite time t_*: $t_0 - (1/H_0) < t_* < t_0$, such that $S(t) \to 0$ as $t \to t_*$; the universe starts at a space-time singularity there, with $\mu \to \infty$ and $T \to \infty$ if $\mu + p/c^2 > 0$.

This is not merely a start to matter — it is a start to space, to time, to physics itself. It is the most dramatic event in the history of the universe: it is the start of existence of everything. The underlying physical feature is the non-linear nature of the EFE: going back into the past, the more the universe contracts, the higher the active gravitational density, causing it to contract even more. The pressure p that one might have hoped would help stave off the collapse makes it even worse because (consequent on the form of the EFE) p enters algebraically into the Raychaudhuri equation (7) with the same sign as the energy density μ. Note that the Hubble constant gives an estimate of the age of the universe: the time $\tau_0 = t_0 - t_*$ since the start of the universe is less than $1/H_0$.

This conclusion can in principle be avoided by a cosmological constant, but in practice this cannot work because we know the universe has expanded by at least a ratio of 6, as we have seen objects at a redshift[6] of 5; from (14), the cosmological constant would have to have an effective magnitude at least $6^3 = 216$ times the present matter density to dominate and cause a turn-around then or at any earlier time, and so would be much bigger than its observed present upper limit (of the

[4] It is however singular in that it is geodesically incomplete; this metric covers only half the de Sitter hyperboloid [Schrödinger, 1956; Hawking and Ellis, 1973].

[5] There is also a static (non-RW) form of the metric — the first form of the metric discovered.

[6] The redshift z for light emitted at t_e and observed at t_0 is related to the expansion by $1 + z = S(t_0)/S(t_e)$, see Sec. 2.3.3.

same order as the present matter density). Accordingly, no turn around is possible while classical physics holds [Ehlers and Rindler, 1989]. However energy-violating matter components such as a scalar field (Sec. 2.6) can avoid this conclusion, if they dominate at early enough times; but this can only be when quantum fields are significant, when the universe was at least 10^{12} smaller than at present.

Because $T_{rad} \propto S^{-1}$ (eqn.(6)), a major conclusion is that *a* Hot Big Bang *must have occurred; densities and temperatures must have risen at least to high enough energies that quantum fields were significant*, at something like the GUT energy. The universe must have reached those extreme temperatures and energies at which classical theory breaks down.

2.2 The hot big bang

The matter and radiation in the universe gets hotter and hotter as we go back in time towards the initial quantum state, because it was compressed into a smaller volume. In this *Hot Big Bang* epoch in the early universe, we can use standard physical laws to examine the processes going on in the expanding mixture of matter and radiation [Weinberg, 1972; Perkins, 2005]. A key feature is that about 300,000 years after the start of the Hot Big Bang epoch, nuclei and electrons combined to form atoms. At earlier times when the temperature was higher, atoms could not exist, as the radiation then had so much energy it disrupted any atoms that tried to form into their constituent parts (nuclei and electrons). Thus at earlier times matter was ionized, consisting of negatively charged electrons moving independently of positively charged atomic nuclei. Under these conditions, the free electrons interact strongly with radiation by Thomson scattering. Consequently matter and radiation were tightly coupled in equilibrium at those times, and the Universe was opaque to radiation. When the temperature dropped through the ionization temperature of about 4000K, atoms formed from the nuclei and electrons, and this scattering ceased: the Universe became very transparent (today we are able to see galaxies at enormous distances from us). The time when this transition took place is known as the *time of decoupling* — it was the time when matter and radiation ceased to be tightly coupled to each other, at a redshift $z_{dec} \simeq 1100$ [Dodelson, 2003]. By (6), the universe was radiation dominated ($\mu_{rad} \gg \mu_{mat}$) at early times and matter dominated ($\mu_{rad} \ll \mu_{mat}$) at late times;[7] matter-radiation density equality occurred significantly before decoupling (the temperature T_{eq} when this equality occurred was $T_{eq} \simeq 10^4$K; at that time the scale factor was $S_{eq} \simeq 10^4 S_0$, where S_0 is the present-day value). The dynamics of both the background model and of perturbations about that model differ significantly before and after S_{eq} [Dodelson, 2003].

[7]The dynamically dominant Cold Dark Matter (Sec. 2.3.6) obeys the same density law (6) as baryons.

2.2.1 Cosmic Blackbody Radiation

Radiation was emitted by matter at the time of decoupling, thereafter travelling freely to us through the intervening space. When it was emitted, it had the form of blackbody radiation, because this is a consequence of matter and radiation being in thermodynamic equilibrium at earlier times. Thus *the matter at* $z = z_{dec}$ *forms the Last Scattering Surface* (LSS) *in the early universe, emitting Cosmic Blackbody Background Radiation*[8] ('CBR') *at 4000K, that since then has travelled freely with its temperature T scaling inversely with the scale function of the universe.*[9] As the radiation travelled towards us, the universe expanded by a factor of about 1100; consequently by the time it reaches us, it has cooled to 2.75 K (that is, about 3 degrees above absolute zero, with a spectrum peaking in the microwave region), and so is extremely hard to observe. It was however detected in 1965, and its spectrum has since been intensively investigated, its blackbody nature being confirmed to high accuracy [Partridge, 1995]. Its existence is now taken as solid proof both that the Universe has indeed expanded from a hot early phase, and that standard physics applied unchanged at that era in the early universe.

2.2.2 Particle interactions and element formation

The thermal capacity of the radiation is hugely greater than that of the matter. At very early times before decoupling, the temperatures of the matter and radiation were the same (because they were in equilibrium with each other), scaling as $1/S(t)$ (eqn.(6)). The early universe exceeded any temperature that can ever be attained on Earth or even in the centre of the Sun; as it dropped towards its present value of 3 K, successive physical reactions took place that determined the nature of the matter we see around us today. At very early times and high temperatures, only elementary particles can survive and even neutrinos had a very small mean free path; as the universe cooled down, neutrinos decoupled from the matter and streamed freely through space. At these times the expansion of the universe was radiation dominated, and we can approximate the universe then by models with $\{k = 0,\ w = 1/3,\ \Lambda = 0\}$, the resulting simple solution of (14) uniquely relating time to temperature:

$$(19)\quad S(t) = S_0\, t^{1/2},\ \ t = 1.92\,\mathrm{sec}\left[\frac{T}{10^{10}K}\right]^{-2}.$$

(There are no free constants in the latter equation).

At very early times, even neutrinos were tightly coupled and in equilibrium with the radiation; they decoupled at about 10^{10}K [Dodelson, 2003, pp. 44-46], resulting in a relic neutrino background density in the universe today of about $\Omega_{\nu 0} \simeq 10^{-5}$ if they are massless (but it could be higher depending on their masses). Key events in the early universe are associated with out of equilibrium phenomena

[8]This is often called "Cosmic Microwave Background", or CMB for short. However it is only microwave at the present epoch.

[9]This scaling for freely propagating radiation follows from the discussion in Sec. 2.3.3.

[Dodelson, 2003, p. 58]. An important event was the era of *nucleosynthesis*, the time when the light elements were formed. Above about 10^9K, nuclei could not exist because the radiation was so energetic that as fast as they formed, they were disrupted into their constituent parts (protons and neutrons). However below this temperature, if particles collided with each other with sufficient energy for nuclear reactions to take place, the resultant nuclei remained intact (the radiation being less energetic than their binding energy and hence unable to disrupt them). Thus the nuclei of the light elements — deuterium, tritium, helium, and lithium — were created by neutron capture. This process ceased when the temperature dropped below about 10^8 K (the nuclear reaction threshold). In this way, the proportions of these light elements at the end of nucleosynthesis were determined; they have remained virtually unchanged since. The rate of reaction was extremely high; all this took place within the first three minutes of the expansion of the Universe. One of the major triumphs of Big Bang theory is that *theory and observation are in excellent agreement provided the density of baryons is low:* $\Omega_{bar0} \simeq 0.044$. *Then the predicted abundances of these elements* (25% *Helium by weight,* 75% *Hydrogen, the others being less than* 1%) *agrees very closely with the observed abundances.* Thus the standard model explains the origin of the light elements in terms of known nuclear reactions taking place in the early Universe [Schramm and Turner, 1998]. However heavier elements cannot form in the time available (about 3 minutes).

In a similar way, physical processes in the very early Universe (before nucleosynthesis) can be invoked to explain the ratio of matter to anti-matter in the present-day Universe: a small excess of matter over anti-matter must be created then in the process of *baryosynthesis*, without which we could not exist today (if there were no such excess, matter and antimatter would have all annihilated to give just radiation [Silk, 2005]). However other quantities (such as electric charge) are believed to have been conserved even in the extreme conditions of the early Universe, so their present values result from given initial conditions at the origin of the Universe, rather than from physical processes taking place as it evolved. In the case of electric charge, the total conserved quantity appears to be zero: after quarks form protons and neutrons at the time of baryosynthesis, there are equal numbers of positively charged protons and negatively charged electrons, so that at the time of decoupling there were just enough electrons to combine with the nuclei and form uncharged atoms (it seems there is no net electrical charge on astronomical bodies such as our galaxy; were this not true, electromagnetic forces would dominate cosmology, rather than gravity).

After decoupling, matter formed large scale structures through gravitational instability [Bothun, 1998, pp. 183-222] which eventually led to the formation of the first generation of stars [Silk, 2005] and is probably associated with the re-ionization of matter [Dodelson, 2003, p. 73]. However at that time planets could not form for a very important reason: there were no heavy elements present in the Universe. The first stars aggregated matter together by gravitational attraction, the matter heating up as it became more and more concentrated, until its temperature exceeded the thermonuclear ignition point and nuclear reactions started burning hydrogen to form helium. Eventually more complex nuclear reactions started in concentric spheres around the centre, leading to a build-up of heavy elements (carbon, nitrogen, oxygen for example), up to iron. These elements can form in stars because there is a long time available (millions of years) for the reactions to take place. Massive stars burn relatively rapidly, and eventually run out of nuclear fuel. The star becomes unstable, and its core rapidly collapses because of gravitational attraction. The consequent rise in temperature blows it apart in a giant explosion, during which time new reactions take place that generate elements heavier than iron; this explosion is seen by us as a Supernova ("New Star") suddenly blazing in the sky, where previously there was just an ordinary star. Such explosions blow into space the heavy elements that had been accumulating in the star's interior, forming vast filaments of dust around the remnant of the star. It is this material that can later be accumulated, during the formation of second generation stars, to form planetary systems around those stars. Thus *the elements of which we are made (the carbon, nitrogen, oxygen and iron nuclei for example) were created in the extreme heat of stellar interiors, and made available for our use by supernova explosions.* Without these explosions, we could not exist.

2.3 Cosmological Observations

Cosmological models only become meaningful when related to astronomical observations [Hoyle, 1960; Sandage, 1961; Ellis, 1971a; Weinberg, 1972]. These are of two kinds: astronomical observations of distant matter tells us what was happening far away in the universe and (because of the finite speed of light) a long time ago. On the other hand observations of nearby objects (matter on Earth, the solar system, nearby stars for example) when related to theories of origins tell us what was happening very near our past world line a very long time ago. The first set of observations may be characterized as "null cone" observations, the second as "geological" observations, one of the most important being the determination of local element abundances, which are then related to nucleosynthesis calculations (Sec. 2.2.2).

Observations are totally dependent on telescope and detector technology [Harwit, 1984; Bothun, 1998]. After the initial establishment of distance scales and the basic evidence of cosmic homogeneity and expansion in the 1920s and 1930s, progress was slow until the 1960s when observations were extended from the optical to the entire electromagnetic spectrum. In recent decades cosmology has changed from a data-poor to a data-rich subject. Massive new data sets are now available because of the extraordinary improvement of telescope, detector, and computer technology in recent decades, particularly the advent of new detectors such as Charge Coupled Devices (CCD's) and fibre optics (enabling simultaneous measurement of hundreds of redshifts). We now have not only optical, ultraviolet, and infrared observations of galaxies, determining luminosities and spectra with unprecedented sensitivity, but also radio, X-ray, and gamma-ray sky surveys. Galaxies have been detected up to a redshift of 6 and we have identified many quasi-stellar objects and gamma-ray bursters as well as multiple images of very distant gravitationally-lensed galaxies [Harwit, 1984]. Large-scale structures (clusters of galaxies, superclusters, walls, and voids) have been identified, with associated large-scale velocity flows [Bothun, 1998, pp. 85-137].

In addition to large-scale number-count and redshift surveys, we have measured the background radiation spectrum and anisotropies at all wavelengths. We identify the radiation as 'background' precisely when it is constant on very small angular scales (as opposed to discrete sources, which appear as isolated objects). There is a complex relation of this radiation to the intergalactic matter density and thermal history. The most important component of the background radiation is the Cosmic Blackbody Radiation ('CBR') mentioned above (Sec. 2.2); detailed observations have mapped its temperature over the whole sky at a sensitivity of better than one part in 10^5. However other components of the background radiation (X-ray and radio in particular) convey important information on the temperature and density of intergalactic matter, and hence strongly restrict its possible thermal history. For example hot matter emits X-rays, so the X-ray background measurement restricts the amount of hot intergalactic matter allowed; while neutral hydrogen strongly absorbs Ultra-Violet radiation to give the Lyman alpha spectral absorption line, so absence of such absorption gives strong limits on the amount of neutral hydrogen and hence on the temperature of intergalactic matter.

2.3.1 Isotropy

The first important point about cosmological observations is that *when averaged on a large enough physical scale* (*clusters of galaxies and above*) *they are statistically isotropic about us*; there is no direction apparently pointing to the centre of the universe. The *high degree of isotropy of the CBR* strongly supports this conclusion: its temperature is the same in all directions about us to better than one part in 10,000 after we have allowed for the motion of the Earth relative to the cosmos (about 250 km/sec), which creates a temperature dipole at one part in a thousand.[10] Any inhomogeneities or anisotropies in the matter distribution lead to anisotropies in this radiation, as recently measured at only one part in 10^5 by the extremely sensitive detectors of the COBE and WMAP satellites. This high degree of isotropy is the major reason we believe the Universe itself is spatially homogeneous and isotropic to a good approximation (see Sec. 4.2.2), providing good observational support for use of the FL universe models as accurate models of the observed region of the universe.

2.3.2 Distance scale and ages

The underlying problem in all astronomy is determining the distances of observed objects. This is done by a 'cosmic distance ladder' [Bothun, 1998, pp. 25-83]

[10]The CBR dipole that could be interpreted as due to a major cosmological inhomogeneity is rather interpreted as being due to our motion ('peculiar velocity') relative to a spatially homogeneous universe.

whereby nearest objects have their distance determined by parallax (i.e. essentially by local trigonometry); and more distant ones by a series of consecutive distance indicators (Cepheid variables, RR Lyrae variables, brightest red supergiants) until at a cosmological distance, redshift z is a primary distance indicator, but is contaminated by local velocities of matter relative to the rest-frame of the universe. Other distance indicators (for example the Tully-Fisher method, the luminosity function of planetary nebulae, the globular cluster luminosity function, surface brightness fluctuations) serve to refine the estimates [Bothun, 1998].

Closely associated with the distance scale is determination of the Hubble constant H_0 (the present rate of expansion of the universe), because estimates of the size of the observable region of the universe scale with the Hubble constant. But the Hubble constant also determines the age of the universe, so its determination underlies a crucial consistency condition for cosmology: *the age of objects in the universe* (*rocks, planets, stars, star clusters, galaxies*) *must be less than the age of the universe.* This condition has been a cause of concern ever since we have had good estimates of ages and of the Hubble constant.[11] It seems not to be violated by current observations of low redshift objects given the current estimates of $H_0 \simeq 70$ km/sec/Mpc, giving an age of the universe of about 15 billion years whereas the oldest star clusters seem to be about 14 billion years old. However it is very tight, perhaps even problematic, for very distant (and so much younger) objects [Jain and Dev, 2005].

2.3.3 Observational relations

Light travels on null geodesics $x^a(\lambda)$ in spacetime (the tangent vector $k^a := dx^a/d\lambda$ is such that $k^a{}_{;b}k^b = 0$, $k^a k_a = 0$). In a RW geometry, it suffices to consider only radial null geodesics (by the symmetries of the model, these are equivalent to generic geodesics). Then from (2) we find that for light emitted at time t_e and received at time t_0, the comoving radial distance $u(t_0, t_1) := r_0 - r_1$ between comoving emitters and receivers is given by

$$(20)\quad \{ds^2 = 0,\ d\theta = 0 = d\phi\} \ \Rightarrow\ u(t_0, t_1) = \int_{t_1}^{t_0} \frac{dt}{S(t)} = \int_{S_1}^{S_0} \frac{dS}{S\dot{S}}$$

with $\dot{S}$ given by the Friedmann equation (9). The key quantities related to cosmological observations are redshift, area distance (or apparent size), and local volume corresponding to some increment in distance (determining number counts) [Sandage, 1961; Ellis, 1971a; Weinberg, 1972]. The redshift z measured for an object emitting light of wavelength λ_e that is observed with wavelength λ_0 is given by

[11] Indeed Hubble himself never fully accepted the expanding universe theory because of age difficulties, preferring to refer to a redshift-distance relation rather than a velocity distance relation [Hubble, 1936]. However the problem has been eased by a series of revisions of the value of the Hubble constant since then, due to a better understanding of the primary distance indicators.

$$(21)\quad 1+z := \frac{\lambda_0}{\lambda_e} = (1+z_c)(1+z_v),$$

where z_v is the redshift caused by the local peculiar motion of the object observed ($z_v = 0$ for comoving objects), and z_c is the cosmological redshift given by

$$(22)\quad 1+z_c = \frac{S(t_0)}{S(t_e)}.$$

From eqn.(21), the same ratio of observed to emitted light holds for all wavelengths: a key identifying property of redshift. The problem in using redshifts of objects as a distance indicator is to separate out the cosmological from the Doppler components, which lead to redshift-space distortions [Dodelson, 2003, pp. 275-282]; this can reasonably be done for a cluster of galaxies by appropriate averaging over cluster members ($\langle z_v \rangle = 0$ for a comoving cluster). The area distance r_0 of an object at redshift z_c and of linear size l which subtends angular size α is given by[12]

$$(23)\quad r_0(z_c) := \frac{l}{\alpha} = f(u)S(t_e).$$

Thus measures of apparent sizes will determine the area distance if the source physical size is known. The flux of radiation F measured from a point source of luminosity L emitting radiation isotropically is given by the fraction of radiant energy emitted by the source in a unit of time that is received by the telescope:

$$(24)\quad F = \frac{L}{4\pi}\frac{1}{f^2(u)S^2(t_0)(1+z)^2};$$

(the two redshift factors account firstly for time dilation observed between observer and source, and secondly for loss of energy due to redshifting of photons). The source's *apparent magnitude* m is defined from the flux: $m = -2.5log_{10}F + const.$ On using (22, 23), equation (24) becomes

$$(25)\quad F = \frac{L}{4\pi}\frac{1}{r_0^2(1+z)^4}.$$

showing that measures of magnitudes will determine the area distance if the source's intrinsic luminosity is known. On using (23) it follows from (25) that *the point-wise surface brightness of extended objects (the flux received per unit solid angle) depends only on redshift* [Kristian and Sachs, 1966; Ellis, 1971a] — a key feature in determining detection probabilities and in gravitational lensing observations. It further follows from this result that *a blackbody spectrum emitted at temperature T_e when observed with a redshift z remains a blackbody spectrum but with observed temperature $T_0 = T_e/(1+z)$* — a crucial feature in analyzing the CBR observations.

[12] This depends only on z_c because apparent shapes and sizes are independent of the motion of the source.

Using the Friedmann equation and the relevant equation of state for matter, the area distance can be determined as a function of redshift z_c in terms of the Hubble constant H_0, deceleration parameter q_0, and cosmological constant Λ. In the case of pressure-free matter with vanishing cosmological constant, one obtains from (20), (9), (22), and (23)[13] the Mattig relation [Sandage, 1961]

$$(26) \quad r_0(z_c) = \frac{1}{H_0 q_0^2} \frac{(q_0 - 1)(1 + 2q_0 z_c)^{1/2} + (q_0(z_c - 1) + 1)}{(1 + z_c)^2} .$$

Consequently measures of either apparent size of sources of known physical size, or of radiant flux from sources of known intrinsic luminosity, will determine the deceleration parameter q_0. Generalizations of this relation hold if a cosmological constant or radiation is present. An interesting aspect is that there is a minimum apparent size for objects of fixed physical size at some redshift $z_c = z_*$ depending on the density parameter and the cosmological constant. The past light cone of the observer attains a maximum area at z_*; the entire universe acts as a gravitational lens for further off objects, magnifying their apparent size so that very distant objects can appear to have the same angular size as nearby ones [Hoyle, 1960; Sandage, 1961; Ellis, 1971a]. For the Einstein-de Sitter universe, the minimum angular diameter is at $z_* = 1.25$; in low density universes, it occurs at higher redshifts.

The number of objects seen in a solid angle $d\Omega$ for a distance increment du (characterized by increments dz, dm in the observable variables z and m) is given by

$$(27) \quad dN = W(u)\,\rho(t_e) S^3(t_e) f(u) du d\Omega$$

where the detection probability or 'selection function' is $W(u)$ [Dodelson, 2003, p. 263] and $\rho(t_e)$ is the number density of objects at the time of emission (spatial homogeneity is expressed in the fact that this is independent of the spatial coordinates). The observed total number N of objects measured in a survey is given by integrating from the observer to the survey limit: in terms of the radial coordinate r_e of the source (which can be related to redshifts or magnitudes), $N = \int_{r_0}^{r_e} dN$. If the number of objects is conserved (e.g. observing galaxies in an epoch when they are neither created nor destroyed), $\rho(t_e) = \rho(t_0)(1 + z)^3$ and we find from (27) that in the idealized case when W is independent of distance (a reasonable assumption for relatively nearby objects),

$$(28) \quad N = W\rho(t_0) d\Omega \int_{r_0}^{r_e} f(u) du.$$

The simple integral has to be separately done for the cases $k = +1, 0, -1$ [Sandage, 1961].

The above equations enable one to determine observational relations between observable variables, for example (m, z), (α, z) or (N, m) relations for objects with known intrinsic properties (known size or luminosity, for example), and so

[13] Or, more elegantly, from the geodesic deviation equation (see [Ellis and van Elst, 1999b]).

to observationally determine q_0. These relations have to be modified if there is absorption by an intergalactic medium, gravitational lensing, or anisotropic emission of radiation; and detailed comparisons with observations have to take into account the spectrum of the source as well as source detection and identification probabilities [Harwit, 1984]. Here we encounter *the contrast between image and reality*: there can be many objects out there that we either do not detect, or do not recognize for what they are [Disney, 1976]. An "observational map" relating source properties to the nature of their images gives a useful view of how this occurs [Ellis *et al.*, 1984].

One important feature here is that *a specific object will look completely different at different wavelengths* (optical, radio, X-ray for example); indeed it may be detectable at one wavelength and not at another. This shows very clearly how our images of reality are dependent on the detectors we use. To get a full picture of what is out there, we need to use multiple modes of investigation — imaging at all wavelengths together with intensity, spectral, and polarization measurements [Harwit, 1984], as well as watching for time variations. A second important feature is observational selection effects such as the *Malmquist bias* — if we have a population of objects with different luminosities, at large distances we will only see the more luminous objects (the fainter ones will not be detected); hence *the average luminosity will appear to increase with distance*, but this is just an observational effect rather than the real state of affairs. Using different detection thresholds controls this effect to some degree.

2.3.4 Number Counts and the visible matter density

Number counts of galaxies as a function of redshift or luminosity show approximate spatial homogeneity of the universe [Hubble, 1936]. However counts of radio sources and quasi-stellar objects (qso's) show that the universe has not been in a steady state as proposed by Bondi, Gold, and Hoyle [Bondi, 1960]. Indeed *number counts are only compatible with a RW geometry if there has been evolution of source numbers and/or luminosities* [Sciama, 1971].

Number counts also give estimates of the density of visible (luminous) matter in the universe: $\Omega_{vm0} \simeq 0.015$. This is very low relative to the critical density ($\Omega_0 = 1$) and is also considerably less than the amount of baryons determined by nucleosynthesis studies ($\Omega_{bar0} \simeq 0.044$, see Sec. 2.2.2). Thus *much of the baryonic matter in the universe is in some hidden (non-luminous) form* [Bothun, 1998, pp. 223-272], e.g. burnt out stars [Hogan, 1999].

2.3.5 Apparent Luminosities and sizes: Dark Energy

Apparent sizes or luminosities as a function of redshift can be used to determine the deceleration parameter q_0 (Sec. 2.1) if the intrinsic source sizes or luminosities are known. The problem is that until recently there were not known enough galaxies or other objects of standard size or luminosity to use to determine q_0, and scatter in their properties leads to biassing of observations by the Malmquist effect. However this dramatically changed with recent observations of the decay curves of the luminosity of supernovae in distant galaxies. It turns out that the peak luminosity of type Ia supernovae is closely correlated with their light curve decay time, for the first time giving reliable 'standard candles' for galaxies at large distances [Perlmutter *et al.*, 1998]. The conclusion from these observations is that, rather than slowing down as expected, *the rate of expansion of the universe is speeding up at a rate corresponding to a cosmological constant with* $\Omega_{\Lambda 0} = 0.7$. This evidence is concordant with that from CBR observations and number counts [Dodelson, 2003; Silk, 2005].

The nature of the field or matter causing this acceleration is unclear. Its equation of state $w := pc^2/\mu$, is unknown, and many physical and unphysical proposals are being made in this regard. From (7), it has to violate the strong energy condition (8) and so must have a large negative pressure. It could indeed be due to a cosmological constant ($w = -1$), which would have dominated the expansion of the universe since a redshift $z \simeq 0.33$, and would have been negligible earlier (and is also negligible on small scales — it does not affect local astrophysics). However it could also be some other form of matter or field with effective negative pressure so that $w < -1/3$, as can happen in the case of a scalar field (see eqn.(33) below). In that case it is called 'quintessence'. There are many speculations as to what this might be, but there is no clarity on the matter. One should note here that alternative explanations of the observations are possible, for they can be exactly reproduced by a spherically symmetric inhomogeneous universe model where we are near the centre [Mustapha *et al.*, 1998], or could at least partly be due to the backreaction of inhomogeneities on the cosmic expansion [Ellis and Buchert, 2006] or the effect of inhomogeneities on the effective area distance [Kantowski *et al.*, 1995; Kantowski, 1998]. These alternatives are being investigated, but the most probable cause remains some unknown kind of matter or field with effective negative energies.

In summary, the standard gravitational equations together with the supernovae observations imply *presence of a cosmological constant or some equivalent form of 'dark energy' with a large effective negative energy density* μ_{grav} *(due to negative pressure) dominating the present expansion of the universe; its physical nature is unknown.* There is no known physics reason why this force should exist at this level where it is just detectable — quantum field theory relates the cosmological constant to the zero-point energy of the vacuum, and suggests it should be enormously larger than observed [Weinberg, 1989; 2000a; 2000b; Rugh and Zinkernagel, 2002; Zinkernagel, 2002; Susskind, 2005]. It is a major mystery why it exists at the small (just detectable) level that observations indicate [Seife, 2003]. A key aspect of present day cosmology is trying on the one hand to observationally determine the effective equation of state of this 'dark energy' (running the field equations backwards to obtain $w(z)$ from the observations [Saini *et al.*, 2000], and in particular determining whether w is constant or varying over time), and on the other attempting to give a plausible theoretical explanation for its physical origin.

2.3.6 Matter Distribution and Motion: Dark Matter

Detailed studies have been made of the distribution of galaxies and their motions. They occur in clusters, in turn making up superclusters imbedded in vast walls surrounding relatively sparsely populated intergalactic voids. The galaxy *luminosity function* characterizes the numbers of galaxies occurring within each luminosity class; the *covariance function* characterizes their spatial clustering [Peebles, 1993b; Dodelson, 2003]. Large scale motions occur for galaxies in clusters, and for the clusters themselves. It is easy to conceive of matter that is hard to detect (for example, small rocks distributed through space); studies of galactic rotation curves and of motions of galaxies in clusters [Bothun, 1998, pp. 139-181] imply *the existence of huge amounts of unseen dark matter, dominating the dynamics of the Universe*: its density is $\Omega_{dm0} \simeq 0.3$, much greater than both visible matter ($\Omega_{vm0} = 0.015$) and baryons ($\Omega_{bar0} = 0.044$), but significantly less than the critical density $\Omega_0 = 1$. Thus *the dark matter is non-baryonic*, meaning it has some kind of exotic nature rather than being the protons and neutrons that are the substance of ordinary matter [Seife, 2003]. In contrast to the 'dark energy' discussed in the previous section, dark matter is dynamically effective on astrophysical scales as well as on cosmological scales. Many attempts have been made to identify its nature, for example it might be axions, supersymmetric partners of known particles, quark nuggets, or massive neutrinos [Gribbin and Rees, 1991; Perkins, 2005], but what it is composed of is still unknown. Laboratory searches are under way to try to detect this matter, so far without success. A key question is whether its apparent existence is due to our using the wrong theory of gravity on these scales. This is under investigation, with various proposals for modified forms of the gravitational equations that might explain the observations without the presence of large quantities of dark matter. This is a theoretical possibility, but the current consensus is that this dark matter does indeed exist.

An important distinction is whether dark matter consists of

(i) weakly interacting massive particles that cooled down quickly, thereafter forming *cold dark matter* ('CDM') moving slowly at the time of structure formation (and resulting in a bottom-up process with large scale structure forming from smaller scale structures), or

(ii) particles that have a low mass and cooled slowly, therefore for a long time forming *hot dark matter*, moving very fast at the time of structure formation (and resulting in a top-down galaxy formation scenario).

Structure formation studies currently support the CDM hypothesis, with hierarchical formation of gravitationally bound objects taking place in a complex bottom up process involving interactions of CDM, baryons, and radiation, with dwarf galaxies forming initially [Silk, 2005; Mouri and Taniguchi, 2005] and then aggregating to form larger structures. These studies are based on massive numerical simulations, with initial conditions provided by the inflationary scenario discussed below, see

Sec. 2.6. Unlike 'dark energy', CDM has an ordinary baryonic equation of state (it is a perfect fluid (4) with $p_{cdm} = 0 \Leftrightarrow w_{cdm} = 0$).

Another way of detecting dark matter in clusters is by its gravitational lensing effects [Schneider *et al.*, 1992]. The bending of light by massive objects was one of the classic predictions of General Relativity theory. Rich clusters of galaxies or galaxy cores can cause strong lensing of more distant objects, where multiple images of distance galaxies and qso's occur, sometimes forming rings or arcs; and weaker lensing by closer masses results in characteristic patterns of distortions of images of distant objects. Analysis of multiple images can be used to reconstruct the lensing mass distributions, and statistical analysis of weak lensing patterns of image distortions are now giving us detailed information on the matter distribution in distant galaxies and clusters. These studies show that to get enough lenses in an almost flat cosmology ($\Omega_0 \simeq 1$) requires the presence of a cosmological constant — there cannot be a critical density of dark matter present [Dodelson, 2003; Silk, 2005].

A key feature of present-day cosmology is attempts to identify the nature of this dark matter, and if possible to detect it in a laboratory situation. While observations favour the CDM scenario, some residual problems as regards the emergence of fine-scale structure still need resolution [Silk, 2005].

2.3.7 The CBR Power spectrum

The CBR angular anisotropies are characterized by an angular power spectrum showing the amount of power in perturbations at each physical scale on the LSS [Bennet et al., 2003; Seife, 2003; Dodelson, 2003]. In the time from the the end of inflation to the LSS, modes of different wavelengths can complete a different number of oscillations. This translates the characteristic periods of the waves into characteristic lengths on the LSS, leading to a series of maxima ('acoustic peaks') and minima in the inhomogeneities on the LSS and consequently in the CBR angular anisotropy power spectrum [Hu and Sugiyama, 1995b; Hu and Sugiyama, 1995a; Peacock, 1999; Perkins, 2005]. These inhomogeneities then form the seeds for structure formation and so are related to the power spectrum of physical scales for structures that form later. They are characterised by a (3-dimensional) spatial power spectrum on the LSS; because we receive the observed CBR radiation from the 2-sphere $S_{2:LSS}$ where our past light cone intersects the LSS, this is seen by us as a 2-dimensional power spectrum of anisotropies on the sky (characterised by the unit sphere S_2 of all direction vectors e_a: $e^a e_a = 1$, $e^a u_a = 0$).

The apparent angular size of the largest CBR peak (about 1^o) allows estimates of the area distance to the LSS and hence of the density of matter in the universe for various values of the cosmological constant, and determines the major cosmological parameters [Spergel *et al.*, 2003]:

> *"By combining WMAP data with other astronomical data sets, we constrain the geometry of the universe:* $\Omega_{tot} = 1.02 \pm 0.02$*, the equation of state of the dark energy,* $w < -0.78$ *(95% confidence limit), and the*

> *energy density in neutrinos, $\Omega_\nu h^2 < 0.0076$ (95% confidence limit). For 3 degenerate neutrino species, this limit implies that their mass is less than 0.23 eV (95% confidence limit). The WMAP detection of early reionization rules out warm dark matter."*

There is however a problem here: while the agreement of theory and observations for all small angular scales is remarkable, there is a divergence at the largest angular scales: the observations show less power than expected. Specifically, the quadrupole and octopole are much lower than theory predicts. Also the axes of the quadrupole and octopole are very precisely aligned with each other, and there are other angular anomalies [Starkman and Schwarz, 2005]. These effects might be due to (i) observational contamination by the galaxy (which gets in the way of our view of the LSS), (ii) a contingent ('chance') event (it represents 'cosmic variance', discussed below, see Sec. 3), (iii) our living in a well-proportioned 'small universe' which is spatially closed so that there is a maximum size to possible fluctuations [Weeks *et al.*, 2003], or (iv) some unexpected new physical effect or deeper problem with our understanding of the early universe. The jury is out as to which the case is; this could turn out to be a crisis for the CBR analysis, but on the other hand one can always just resort to saying it is a statistical fluke (the underlying problem here being the uniqueness of the universe, as discussed in Sec. 3).

There are similar expected peaks in the polarization spectrum of this radiation, and polarization maps should have a mode associated with gravitational waves predicted by inflation to exist in the very early universe (Sec. 2.6); detection of such modes will be a crucial test of inflation [Dodelson, 2003; Sievers *et al.*, 2005b]. Studies of polarization indicate that reionisation of the universe took place as early as a redshift of 17, contrary to what is deduced from qso studies. More detailed studies of anisotropies involve the Sunyaev-Zel'dovich effect (changes in the observed temperature due to scattering by hot matter in galaxy clusters) and gravitational lensing.

There is a huge amount of information in the CBR maps, and their more accurate measurement and interpretation is a central feature of current cosmology [Steinhardt, 1995; Peacock, 1999; Dodelson, 2003; Perkins, 2005].

2.4 Causal and visual horizons

A fundamental feature affecting the formation of structure and our observational situation is the limits arising because causal influences cannot propagate at speeds greater than the speed of light. Thus the region that can causally influence us is bounded by our past null cone. Combined with the finite age of the universe, this leads to the existence of particle horizons limiting the part of the universe with which we can have had causal connection.[14]

A *particle horizon* is by definition comprised by the limiting worldlines of the furthest matter that ever intersects our past null cone [Rindler, 1956; 2001]. This

[14]There are also *event horizons* and *apparent horizons* in some cosmological models [Rindler, 1956; Tipler *et al.*, 1980] and [Rindler, 2001, pp. 376-383].

is the limit of matter that can have had any kind of causal contact with since the start of the universe, characterized by the comoving radial coordinate value

$$(29)\quad u_{ph} = \int_0^{t_0} \frac{dt}{S(t)}\,.$$

The present physical distance to the matter comprising the horizon is

$$(30)\quad d_{ph} = S(t_0)u_{ph}.$$

The key question is whether the integral (29) converges or diverges as we go to the limit of the initial singularity where $S \to 0$. *Horizons will exist in standard FL cosmologies for all ordinary matter and radiation*, for u_{ph} will be finite in those cases; for example in the Einstein-de Sitter universe (see Sec. 2.1.1), $u_{ph} = 3t_0^{1/3}$, $d_{ph} = 3t_0$. We will then have seen only a fraction of what exists, unless we live in a universe with spatially compact sections so small that light has indeed had time to traverse the whole universe since its start; this will not be true for universes with the standard simply-connected topology. Penrose's powerful use of conformal methods (see [Hawking and Ellis, 1973; Tipler *et al.*, 1980]) gives a very clear geometrical picture of the nature of horizons [Ellis and Williams, 2000]. They may not exist in non-FL universes, for example Bianchi (anisotropic) models [Misner, 1969]. In universes with closed spatial sections, a supplementary question arises: Is the scale of closure smaller than the horizon scale? There may be a finite time when causal connectivity is attained, and particle horizons cease to exist. In standard $k = +1$ FL models, this occurs just as the universe reaches the final singularity; if however there is a positive cosmological constant or other effective positive energy density field, it will occur earlier. The horizon always grows, because (29) shows that u_{ph} is a monotonically increasing function of t_0. Despite many contrary statements in the literature, *it is not possible that matter leave the horizon once it has entered.* In a (perturbed) FL model, once causal contact has taken place, it remains until the end of the universe.

The importance of horizons is two-fold: they underlie causal limitations relevant in the origin of structure and uniformity [Misner, 1969; Guth, 1981], and they represent absolute limits on what is testable in the universe [Ellis, 1975; 1980].

2.4.1 Causal limitations

As to causal limitations, horizons are important in regard both to the smoothness of the universe on large scales, and the lumpiness of the universe on small scales. The issue of smoothness is encapsulated in the *horizon problem*: if we measure the temperature of the CBR arriving here from opposite directions in the sky in a standard FL model, it came from regions of the surface of last scattering that can have had no causal contact of any kind with each other since the start of the universe. In a radiation-dominated early universe with scale factor (19), the size of the particle horizon at the time of last scattering appears as an angular scale of about 1^o in the sky today, and corresponds to a comoving physical length of about

400,000 light years when evaluated today. Why then are conditions so similar in these widely separated regions? [Misner, 1968; Guth, 1981; Blau and Guth, 1987; Kolb and Turner, 1990]. Note that this question is of a philosophical rather than physical nature, i.e. there is no contradiction here with any experiment, but rather an unease with an apparent fine tuning in initial conditions. This problem is claimed to be solved by the inflationary universe scenario mentioned below, see Sec. 2.6.

Associated with the existence of horizons is the prediction that physical fields in different regions in the universe should be uncorrelated after symmetry breaking takes place, because they cannot have interacted causally. Consequently, if grand unified theories are correct, topological defects such as monopoles and cosmic strings may be expected as relics of the expansion of the very early universe [Kolb and Turner, 1990]. In a standard cosmology, far too many monopoles are predicted. This is also solved by inflation.

As to the lumpiness, the issue here is that if we believe there was a state of the universe that was very smooth — as indicated at the time of decoupling, by the low degree of anisotropy of the CBR, and represented by the RW geometry of the FL models — then there are limits to the sizes of structures that can have grown since then by causal physical processes, and to the relative velocities of motion that can have been caused by gravitational attraction in the available time (for example, the peculiar motion of our own galaxy relative to the CBR rest frame caused by the huge overdensity called the 'Great Attractor'). If there are larger scale structures or higher velocities, these must have been imprinted in the perturbations at the time of last scattering, for they cannot have been generated in a causal way since that time. They are set into the initial conditions, rather than having arisen by physical causation from a more uniform situation. This is a key factor in the theory of growth of perturbations in the early universe where the expansion damps their growth. The quantity determining the relevant physical scales for local causal influences in an expanding universe is the *comoving Hubble radius* $\lambda_H := (SH)^{-1}$; the way perturbations of wavelength λ develop depends on whether $\lambda > \lambda_H$ or $\lambda < \lambda_H$ [Dodelson, 2003, pp. 146-150].

Actually the domain of causal influence is even more tightly constricted than indicated by the past light cone: the limits coming from the horizon size are limits on what can be influenced by particles and forces acting at the speed of light. However only freely travelling photons, massless neutrinos, and gravitons can move at that speed; and such particles coming from cosmological distances have very little influence on our galaxy or the solar system (indeed we need very delicate experiments to detect them). Any massive particles, or massless particles that are interacting with matter, will travel slower (for example before decoupling, light has a very small mean free path and information will travel only by sound waves and diffusion in the tightly coupled matter-radiation fluid). The characteristics for pressure-free scalar and vector perturbations are timelike curves, moving at zero velocity relative to the matter; while density perturbations with pressure can move at the speed of sound, only tensor perturbations can travel at the speed of light. Thus the true domain that influences us significantly is much less than indicated by the particle horizon. It is the small region round our past world line characterised after decoupling by the comoving scale from which matter coalesced into our galaxy: a present distance of about 1 to 1.95 Mpc,[15] corresponding to an observed angle of about 0.64 arcminutes on the LSS. Before decoupling it would have been limited by the sound horizon [Dodelson, 2003, p. 257] rather than the particle horizon.

2.4.2 Observational limitations

Clearly we cannot obtain any observational data on what is happening beyond the particle horizon; indeed we cannot even see that far because the universe was opaque before decoupling. *Our view of the universe is limited by the* visual horizon, *comprised of the worldlines of furthest matter we can observe — namely, the matter that emitted the CBR at the time of last scattering* [Ellis and Stoeger, 1988]. This occurred at the time of decoupling $t = t_{dec}$ ($z_{dec} \simeq 1100$), and so the visual horizon is characterized by $r = u_{vh}$ where

$$(31)\quad u_{vh} = \int_{t_{dec}}^{t_0} \frac{dt}{S(t)} < u_{ph}.$$

Indeed the LSS delineates our visual horizon in two ways: we are unable to see to *earlier times* than its occurrence (because the early universe was opaque for $t < t_{dec}$), and we are unable to detect matter at *larger distances* than that we see on the LSS (we cannot receive radiation from matter at co-moving coordinate values $r > u_{vh}$). The picture we obtain of the LSS by measuring the CBR from satellites such as COBE and WMAP is just a view of the matter comprising the visual horizon, viewed by us at the time in the far distant past when it decoupled from radiation. The position of the visual horizon is determined by the geometry since decoupling. Visual horizons do indeed exist, unless we live in a small universe, spatially closed with the closure scale so small that we can have seen right around the universe since decoupling. This is a possibility that will be discussed below (Sec. 4.3.1). There is no change in these visual horizons if there was an early inflationary period, for inflation does not affect the expansion or null geodesics during this later period. The major consequence of the existence of visual horizons is that many present-day speculations about the super-horizon structure of the universe — e.g. the chaotic inflationary theory (Sec. 2.6) — are not observationally testable, because one can obtain no definite information whatever about what lies beyond the visual horizon [Ellis, 1975; 1980]. This is one of the major limits to be taken into account in our attempts to test the veracity of cosmological models (Sec. 4.3).

[15]Dodelson [Dodelson, 2003], p. 283; W Stoeger, private communication.

2.5 Theoretical Developments

The cosmological application of Einstein's Theory of Gravitation has also progressed greatly in past decades, as regards exact solutions and generic properties of the field equations; as regards approximate solutions; and in terms of understanding the relationship between them.

2.5.1 Exact solutions and generic properties

Theory initially predicted there must have been a start to the universe, but it was not clear for a long time if this was simply due to the very special exactly isotropic and spatially homogeneous geometry of the standard Friedmann-Lemaître models. It was possible that more realistic models with rotation and acceleration might show the prediction was a mathematical artefact resulting from the idealized models used. The singularity theorems developed by Penrose and Hawking [Hawking and Ellis, 1973; Tipler *et al.*, 1980; Earman, 1999] showed this was not the case: even for realistic geometries, classical gravitational theory predicts a beginning to the universe at a space-time singularity, provided the usual energy conditions were satisfied. This study has led inter alia to a greatly increased understanding of causality and topology of generic universe models [Tipler *et al.*, 1980], including the fact that singularities may have a quite different nature than those in the Robertson-Walker models, for example being anisotropic [Tipler *et al.*, 1980] or of a non-scalar nature [Ellis and King, 1974].

Various classes of exact cosmological solutions are known (Kantowski-Sachs and Bianchi spatially homogeneous but anisotropic models, Tolman-Bondi spherically symmetric inhomogeneous models, and 'Swiss-Cheese' non-analytic models) enabling understanding of dynamical and observational behaviour of more general classes of models than just the FL models [Ellis and van Elst, 1999a]. Dynamical systems studies [Wainwright and Ellis, 1996; Uggla *et al.*, 2003] relate the behaviour of whole classes of anisotropic models in suitable state spaces, enabling identification of generic patterns of behaviour (fixed points, saddle points, attractors, etc.) and hence the relationship between dynamics of higher symmetry and lower symmetry universes. These studies help understanding to what degree the FL models are generic within the families of possible cosmological models, and which models might give observations similar to those in the FL models. In particular they are relevant in considering the possible geometry of the universe at very early or very late times.

2.5.2 Perturbation theory, the gauge issue, and back reaction

Sophisticated perturbation theory has been developed to underlie the theory of structure formation in the expanding universe, examining *the dynamics of perturbed FL models*. The fluid flow in these models can have shear, vorticity, and acceleration, and the Weyl tensor C_{ijkl} (see (1)) is not zero, so that density variations, tidal forces, peculiar velocities, and gravitational waves can be present. Detailed studies use the kinetic theory approximation for matter (electrons, protons, dark matter) and radiation (photons, neutrinos), with their dynamics described by the Boltzmann equation [Dodelson, 2003, Ch. 4]; [Uffink, 2006], interacting with space-time inhomogeneities characterised by a perturbed FL metric. A key issue here is the *gauge problem* — how to choose the background model in the perturbed spacetime [Ellis and Stoeger, 1987]. If this is not properly handled then one may attain apparent perturbation solutions that are pure gauge (they are mathematical rather than physical), so that one can alter the apparent growth rate simply by changing coordinates. The key to handling this is either to keep careful track at all stages of remaining gauge freedom and possible changes of gauge, or (preferably, in my view) to use gauge invariant variables (see [Bardeen, 1980; Ellis and Bruni, 1989; Challinor and Lasenby, 1998]).

Most of the literature on perturbation theory deals with the linear case, but some studies tackle the non-linear regime (e.g. [Langlois and Vernizzi, 2005]), and some consider questions such as the origin of magnetic fields and the causes of galactic rotation. A key problem here is properly relating relativistic analyses of astrophysical dynamics to the Newtonian approaches most often used by astrophysicists (e.g. [Bothun, 1998], pp. 183-222); this is not straightforward.[16] A further unresolved issue is the nature of gravitational entropy [Penrose, 1989b; Ellis, 2002; Penrose, 2004]. Many statements about the nature of entropy in physics textbooks are wrong when gravity is dominant, leading to the spontaneous formation of structures such as stars and galaxies. There is as yet no agreed definition of gravitational entropy that is generally applicable; until there is, cosmological arguments relying on entropy concepts are ill-founded.

The existence of inhomogeneities in the universe raises the issue of fitting and back-reaction. To what degree does the nature of the exactly smooth FL models reflect the geometrical and dynamical nature of more realistic ‘lumpy’ universe models? [Ellis and Stoeger, 1987]. Inhomogeneities lead to extra terms appearing in the evolution equations for the idealized background models, representing the back-reaction of the perturbations on their dynamics [Ellis, 1984]. These could possibly be dynamically significant [Ellis and Buchert, 2006], but this is a matter of dispute.

2.6 Inflation

Particle physics processes dominated the very early eras, when exotic processes took place such as the condensation of a quark-gluon plasma to produce baryons. Quantum field theory effects were significant then, and this leads to an important possibility: scalar fields producing repulsive gravitational effects could have

[16] Some exact General Relativity results, which must necessarily apply in the Newtonian limit of General Relativity, have no Newtonian analogue; an example is the shear-free theorem applying to pressure-free matter [Ellis, 1967]. The underlying issue is that there are 10 field equations to be satisfied in General Relativity, with 20 integrability conditions (the Bianchi identities), but only one field equation to be satisfied in Newtonian theory (Poisson’s equation) together with 4 conservation equations.

dominated the dynamics of the universe at those times. This leads to the theory of the inflationary universe, proposed by Alan Guth [1981; 1997]: if $\mu_{grav} = \mu + 3p/c^2 < 0$, which can happen if a scalar field dominates the dynamics of the early universe, an extremely short period of accelerating expansion will precede the hot big bang era [Blau and Guth, 1987]. This produces a very cold and smooth vacuum-dominated state, and ends in 'reheating': conversion of the scalar field to radiation, initiating the hot big bang epoch. This inflationary process is claimed to explain the puzzles mentioned above (Sec. 2.4.1): why the universe is so special (with spatially homogeneous and isotropic geometry and a very uniform distribution of matter), and also why the space sections are so close to being flat at present (we still do not know the sign of the spatial curvature), which requires very fine tuning of initial conditions at very early times. Inflationary expansion explains these features because particle horizons in inflationary FL models will be much larger than in the standard models with ordinary matter, allowing causal connection of matter on scales larger than the visual horizon, and inflation also will sweep topological defects outside the visible domain.

In more detail: in the case of a single scalar field ϕ with spacelike surfaces of constant density, on choosing u^a orthogonal to these surfaces, the stress tensor has a perfect fluid form with

$$(32)\quad \mu = \frac{1}{2}\dot{\phi}^2 + V(\phi), \quad p/c^2 = \frac{1}{2}\dot{\phi}^2 - V(\phi),$$

and so

$$(33)\quad \mu + 3p/c^2 = 2\dot{\phi}^2 - 2V(\phi).$$

The slow-rolling case is $\dot{\phi}^2 \ll V(\phi)$, leading to $\mu + p/c^2 = 2\dot{\phi}^2 \simeq 0 \Rightarrow \mu + 3p/c^2 \simeq -2\mu < 0$. This then enables a resolution of the horizon problem in inflationary FL models: if sufficient inflation took place in the early universe, then all the regions from which we receive CBR were causally connected; indeed if the universe began in an inflationary state, or was inflationary with compact spatial sections, there may be no causal horizons at all. The inflationary models also cause initial perturbations to die away, including velocity perturbations, hence explaining the observed smoothness of the universe on large scales. This process is expected to create a universe with spatially very flat sections at late times:

$$(34)\quad \Omega_0 = \Omega_{dm\,0} + \Omega_{\Lambda\,0} \simeq 1 \Leftrightarrow \Omega_k \simeq 0.$$

This theory led to a major bonus: a proposal that initial tiny quantum fluctuations were expanded to such a large scale by inflation that they provided seeds initiating growth by gravitational attraction of large scale structures such as clusters of galaxies. This theory makes clear observational predictions for the spectrum of CBR anisotropies, which have since been spectacularly verified by observations from balloons and satellites, such as WMAP [Spergel *et al.*, 2003]. Thus inflation has provided us with our first coherent theory of structure formation. Inhomogeneities started as quantum fluctuations in the inflationary epoch which are then amplified in physical scale by the inflationary expansion but remain constant in amplitude when larger than the contemporary Hubble scale, leading to Gaussian scale-free perturbations at the start of the HBB era. Starting from these fluctuations, Cold Dark Matter ('CDM') creates potential wells for baryons to fall into, but the radiation (tightly coupled to the electrons and baryons) resists collapse. Gravity wins if the wavelength λ is greater than the *Jean's length* λ_J (which is proportional to the speed of sound [Rees, 1995; Ellis and van Elst, 1999a]). There are acoustic oscillations (sound waves) when $\lambda < \lambda_J$; these oscillations ceased at decoupling, which led to a dramatic decrease in λ_J and the growth of structure by gravitational instability in a 'bottom up' way (Sec. 2.3.6).[17]

A popular version of inflation is *chaotic inflation* [Linde, 1990; Guth, 2001; Susskind, 2005] where inflation ends at different times in different places, so that one ends up with numerous 'pocket universe' (expanding universe domains like the one we see around us, or perhaps very different) all imbedded in a still-inflating universe region and starting at different times, the whole forming a fractal-like structure. It is argued this is an inevitable consequence of the nature of plausible scalar field potentials.

Inflation is not an inevitable conclusion, for there are some alternatives proposed [Hollands and Wald, 2002; Khoury *et al.*, 2001], and the WMAP results can be reproduced by any scenario where Gaussian scale-free perturbations of suitable amplitude occur at the start of the Hot Big Bang era. However inflation is regarded by most cosmologists as the best proposal available for the era prior to the Hot Big Bang epoch, leading to the presence of such perturbations. Nevertheless one should note it is a generic proposal for what happened, rather than a specific physical theory. While a great many possibilities have been proposed (it could for example be an effective field due to higher-order gravity effects, or it could involve multiple scalar fields), at the present time the identity of the proposed inflationary field ('the inflaton') has not been established or linked to any known particle or field. The hoped-for link between early universe dynamics and particle physics is potential rather than real [Earman and Mosterin, 1999]. Detailed studies of the CBR anisotropies and structure formation in conjunction with the observations hope to distinguish between the various possibilities, for example testing whether the spectral index n takes the scale-free value: $n = 1$, or whether rather there is a tilted power spectrum ($n \neq 1$). A unique spectrum of gravitational waves will also be produced at very early times in an inflationary universe, and detection of these waves either directly by proposed gravitational wave detectors or indirectly by measuring the associated curl mode in the CBR polarization will be an important test of inflation, for example determining the ratio r of scalar to tensor perturbations in the early universe [Dodelson, 2003].

[17] This is a highly simplified account; for more detailed versions, see e.g. [Dodelson, 2003; Silk, 2005].

2.7 *The very early universe*

Quantum gravity processes are presumed to have dominated the very earliest times, preceding inflation. There are many theories of the quantum origin of the universe, but none has attained dominance. The problem is that we do not have a good theory of quantum gravity [Rovelli, 2006], so all these attempts are essentially different proposals for extrapolating known physics into the unknown. A key issue is whether quantum effects can remove the initial singularity and make possible universes without a beginning. Preliminary results suggest this may be so [Bojowald, 2001; Rovelli, 2004; Mulryne *et al.*, 2005].

2.7.1 *Is there a quantum gravity epoch?*

A preliminary issue is, can there be a non-singular start to the inflationary era, thus avoiding the need to contemplate a preceding quantum gravity epoch? In the inflationary epoch the existence of an effective scalar field leads to a violation of the strong energy condition (8), therefore at first sight it seems that a bounce may be possible preceding the start of the expanding inflationary era and avoiding the inevitability of a quantum gravity epoch.

However a series of theorems suggest that inflationary models cannot bounce: they are stated to be future infinite but not past infinite [Guth, 2001]. This is an important issue, so it is worth looking at it further. There are two major requirements to get a bounce. The Friedmann equation (9) relates the scale factor $S(t)$, curvature constant k, and the effective total energy density $\mu(t)$, which is *defined* by this equation whatever dynamics may be involved (multiple scalar fields, higher order gravity, higher dimensional theories leading to effective 4-dimensional theories, etc.).[18] The Raychaudhuri equation (7) includes the effective total pressure $p(t)$, which again is *defined* by this equation. In this section, a cosmological constant Λ is represented as perfect fluid with $\mu_\Lambda + p_\Lambda/c^2 = 0$. To get a bounce, first one needs the curve $S(t)$ of the scale factor as a function of time to bend up: that is,

$$(35)\quad \frac{\ddot{S}}{S} \geq 0 \Leftrightarrow \mu + 3p/c^2 < 0,$$

which is just a violation of the strong energy condition (8). This is the case if $\mu + p/c^2 = 0$ (a vacuum); and indeed by eqn.(33) it is possible for example for any slow-rolling scalar field. Second, one also needs a time when the scale factor is a minimum. Thus there must be a time t_* such that $\dot{S}(t_*) = 0$. From the Friedmann equation (9),

$$(36)\quad \dot{S}^2(t_*) = 0 \Leftrightarrow \frac{\kappa\mu(t_*)}{3} = \frac{k}{S^2(t_*)}.$$

[18]See [Copeland *et al.*, 2005] for the ways various quantum gravity theories result in modified Friedmann equations.

With $k \leq 0$ this is possible only if $\mu(t_*) < 0$. Even with a scalar field (see eqn.(32)) this can only be achieved by having negative potential energies, which appears to be an unphysical requirement. With $k = +1$ this is possible with $\mu(t_*) > 0$ [Robertson, 1933], which is compatible with ordinary matter.

Thus if you want a bounce in an inflationary universe, it is sensible to look to $k = +1$ inflationary models, which indeed will turn around if a vacuum domain occurs for long enough (curvature will eventually always win over a vacuum as we go back into the past [Ellis *et al.*, 2002b; Ellis *et al.*, 2002a]). The theorems mentioned above do *not* include this case (see [Guth, 2001]); they only consider inflationary universes with $k = 0$ and $k = -1$. And one should note here that although the scale-free $k = 0$ exponential case clearly is the model underlying the way many people approach the problem, it is highly exceptional — it is of zero measure within the space of all inflationary FL models.

Explicit non-singular models can be constructed, the simplest being the de Sitter universe in the $k = +1$ frame (Sec. 2.1.1), which is an exact eternal solution that bounces at a minimum radius S_0. This model has the problem that it does not exit inflation (it corresponds to an exactly constant potential), but variants exist where exit is possible; there are also viable non-singular models that start off asymptotic to the Einstein Static universe in the distant past and avoid the need for a quantum gravity epoch [Ellis and Maartens, 2004]. These models start off in a very special state, precisely because they are asymptotic to the Einstein static universe in the distant past. This is a possible situation. It seems likely that the options for the start of inflation are (i) avoiding the quantum gravity era, but at the cost of having special ('fine tuned') initial conditions, or (ii) having a quantum gravity epoch preceding the inflationary era. Thus a key issue is whether the start of the universe was very special or generic.

2.7.2 Quantum gravity effects: The origin of the universe

Contemporary efforts to explain the beginning of the universe, and the particular initial conditions that have shaped its evolution, usually adopt some approach or other to applying quantum theory to the creation of the universe [Lemaître, 1931]. Many innovative attempts have been made here; as this article focuses on General Relativity and its application to cosmology, and it would be impossible to do justice to the various approaches to quantum cosmology [Rovelli, 2006] without a very much longer article. I will just make a few comments on these approaches.

The attempt to develop a fully adequate quantum gravity approach to cosmology is of course hampered by the lack of a fully adequate theory of quantum gravity, as well as by the problems at the foundation of quantum theory (the measurement problem, collapse of the wave function, etc., see [Isham, 1997; Dickson, 2006; Landsman, 2006]) which can be ignored in many laboratory situations but have to be faced in the cosmological context [Perez *et al.*, 2005]. The various attempts at quantum cosmology each develop in depth some specific aspect of quantum theory that may be expected to emerge from a successful theory of quantum gravity applied to the universe as a whole, being for example based on either (i) the Wheeler-DeWitt equation and the idea of the wave function of the universe, or (ii) on some version of embedding in higher dimensional space time (inspired by string theory), or (iii) an appropriate application of loop quantum gravity. In effect they attempt either

(a) to give a true theory of creation *ex nihilo* [Vilenkin, 1982]; such efforts however cannot truly "solve" the issue of creation, for they rely on some structures or other (e.g. the elaborate framework of quantum field theory and much of the standard model of particle physics) pre-existing the origin of the universe, and hence themselves requiring explanation; or

(b) to describe a self-sustaining or self-referential universe which by-passes the issue of creation, either by

    (b1) originating from an eternally pre-existing state, via the recurring idea of a Phoenix universe [Dicke and Peebles, 1979] (as in Veneziano's 'pre-big bang theory' based on analogues of the dualities of string theory, or self-repeating universes such as the chaotic inflationary models of Linde); creation from fluctuations in some quite different pre-existing structure (e.g. emergence from de Sitter space time; or the 'ekpyrotic universe' initiated by a collision between pre-existing 'branes' in a higher dimensional spacetime); or emerging from an eternal static initial state; or

    (b2) starting from a state with different properties of time than usual (or with an an emergent notion of time): as in the Hartle–Hawking no-boundary proposal [Hawking, 1987; Hawking, 1993], and the Gott causal violation proposal [Gott and Li, 1998] where the universe 'creates itself' and starts normal expansion in the domain without closed timelike lines.

Any of these may be combined with one or other proposals for

(c) an effective ensemble of universes [Tegmark, 2003], realized either

    (c1) in space-time regions that are part of either a larger entangled quantum entity, or are part of a single classical space-time, but are effectively disconnected from each other, or

    (c2) in truly disconnected form.

All of these proposals however are strongly speculative, none being based solidly in well-founded and tested physics, and none being in any serious sense supported by observational evidence. They are all vast extrapolations from the known to the unknown. They may or may not be true. One thing is certain: they can't all be true!

2.8 The concordance model

Observational support for the idea of expansion from a Hot Big Bang epoch is very strong, the linear magnitude-redshift relation for galaxies demonstrating the expansion,[19] with source number counts and the existence of the blackbody CBR being strong evidence that there was indeed evolution from a hot early stage. Agreement between measured light element abundances and the theory of nucleosynthesis in the early universe confirms this interpretation. This basic theory is robust to critical probing. Much present activity attempts to link particle physics interactions during very early stages of the expansion of the universe to the creation of structures by gravitational instability much later, traces of the early seed fluctuations being accessible to us through present day CBR anisotropy patterns. Thus the present dominant cosmological paradigm is *a quantum gravity era of some kind followed by inflation; a hot big bang epoch; decoupling of matter and radiation; and then gravitational instability leading to formation of clusters of galaxies from the seed density perturbations that occur on the LSS.*

Together with supernova data, analysis of the CBR angular anisotropies and in particular their peaks gives a *concordance model* of this kind [Bennet et al., 2003; Tegmark, 2002; Tegmark *et al.*, 2004; Dodelson, 2003; Scott, 2005] that is then confirmed by the statistics of matter clustering [Eisenstein *et al.*, 2005a] together with observations of gravitational lensing and large-scale motions of matter [Silk, 2005]. This model is characterized by specific values for a set of cosmological parameters [Liddle, 2004], in particular

(37) $\Omega_{cdm0} \simeq 0.3$, $\Omega_{\Lambda 0} \simeq 0.7$, $T_{cbr0} = 2.75K$, $H_0 \simeq 65$km/sec/mpc, $t_0 \simeq 1.4 \times 10^{10}$years.

Also $\Omega_{bar0} \simeq 0.044$ is the density of baryons, $\Omega_{vis0} \simeq 0.015$ that of luminous matter, and $\Omega_{\nu 0} \simeq 10^{-5}$ that of massless neutrinos, implying $\Omega_0 \simeq 0.3 + 0.7 \simeq 1$ in agreement with the inflationary prediction (34). The sign of k is uncertain, but if the combined evidence of all current observations is taken at face value it is positive, with $\Omega_0 = 1.02 \pm 0.02$ [Spergel *et al.*, 2003]. As noted above, there are some concerns firstly over age issues (see Sec. 2.3.2); secondly concerning the large angle CBR anisotropies (see Sec. 2.3.7); and thirdly regarding details of CDM structure formation at small scales (see Sec. 2.3.6); but none of these issues seems to be crucial at present.

2.8.1 Some misunderstandings

Despite its simplicity, there are some common misconceptions about the standard universe models (cf. [Lineweaver and Davis, 2005]) that can lead to philosophical misunderstandings.

Misconception 1: *The universe is expanding into something.* It is not, as it is all there is. It is just getting bigger, while always remaining all that is. One

[19]The alternative interpretation as gravitational redshifts in a static universe does not work because of the linearity of the observed redshift-distance relation [Ellis *et al.*, 1978].

should note here that a RW universe can be represented as a 4-dimensional curved spacetime expanding in a 5-dimensional flat embedding space time [Robertson, 1933]; however there is no necessity to view the 5-dimensional spacetime in this representation as physically real. Furthermore this embedding is no longer possible when we take perturbations into account; a 10 dimensional flat spacetime is needed for locally embedding a realistic (perturbed) universe model (and to do so globally requires many more dimensions, in general).

Misconception 2: *The universe expands from a specific point, which is the centre of the expansion.* All spatial points are equivalent in these universes, and the universe expands equally about all of them. Every fundamental observer sees exactly the same thing in an exact RW geometry. There is no centre to a FL universe.

Misconception 3: *Matter cannot recede from us faster than light.* It can, at an instant; two distantly separated fundamental observers in a surface $\{t = const\}$ can have a relative velocity greater than c if their spatial separation is large enough [Rothman and Ellis, 1993; Davis and Lineweaver, 2004]. No violation of special relativity is implied, as this is not a local velocity difference, and no information is transferred between distant galaxies moving apart at these speeds. For example, there is presently a sphere around us of matter receding from us at the speed of light;[20] matter beyond this sphere is moving away from us at a speed greater than the speed of light. The matter that emitted the CBR was moving away from us at a speed of about $61c$ when it did so [Rothman and Ellis, 1993].

Misconception 4: *The existence of a preferred RW frame (that in which the universe appears isotropic) contradicts relativity theory, which says all reference frames are equally good.* But this equivalence of frames is true for the equations rather than their solutions. Almost all particular solutions will have preferred world lines and surfaces; this is just a particular example of a *broken symmetry* — the occurrence of solutions of equations with less symmetries than the equations display. This feature is a key theme in modern physics [Brading and Castellani, 2006; Harvey, 2006].

Misconception 5: *The space sections are necessarily infinite if $k = 0$ or -1.* This is only true if they have their 'natural' simply connected topology. If their topology is more complex (e.g. a 3-torus) they can be spatially finite [Ellis, 1971a; Lachièze *et al.*, 1995]. There are many ways this can happen; indeed if $k = -1$ there is an infinite number of possibilities.

Misconception 6: *Inflation implies spatial flatness* $(k = 0 \Leftrightarrow \Omega_k = 1)$ *exactly.* There is nothing in inflationary theory which determines the sign of the spatial curvature. Inflationary universes are very nearly flat at late times; this is very different from being exactly flat (a condition which requires *infinite* fine tuning of initial conditions; if say the two millionth digit in the value of Ω_k is non-zero at

[20] This sphere is not the same as the particle horizon, as is sometimes claimed (see [Rothman and Ellis, 1993]).

any time, then the universe is not spatially flat). Inflationary theory does not have the theoretical teeth required to imply that the universe has exactly flat spatial sections; hence a key issue for cosmology is observationally determining the sign of the spatial curvature, which is potentially dynamically important in both the very early universe [Ellis *et al.*, 2002b; Ellis *et al.*, 2002a] and the late universe (it determines if recollapse is possible, should the dark energy decay away).

2.8.2 Overall

Cosmology has changed from a speculative enterprize into a data-driven science that is part of standard physical theory [Barnett *et al.*, 1996]; a wealth of observations supports this dominant theory [Peebles *et al.*, 1991; Silk, 1997; Perkins, 2005]. Nevertheless some theoretical proposals are being made for the very early stages that have no observational support; and sometimes it may be impossible to ever obtain such support, both as regards the proposed physics and the geometry. Thus in some respects it remains a principle driven enterprise, with observation subordinate to theory.

We now explore the relation between cosmology and philosophy in terms of a series of *Theses* clustered around a set of major *Issues*. One can obtain a synoptic overview of the overall argument by simply considering the full set of *Issues* and *Theses*. They are summarized in the Table at the end.

3 ISSUE A: THE UNIQUENESS OF THE UNIVERSE.

The first and most fundamental issue is that there is only one Universe [Munitz, 1962; McCrea, 1960; Ellis, 1991]. This essential uniqueness of its object of study sets cosmology apart from all other sciences. In particular, the unique initial conditions that led to the particular state of the universe we see were somehow "set" by the time that physical laws as we know them started governing the evolution of both the universe and its contents, whenever that time may be. We cannot alter these unique initial conditions in any way — they are given to us as absolute and unchangeable, even though they are understood as contingent rather than necessary; that is, they could have been different while still being consistent with all known physical laws. The implications are that

Thesis A1: The universe itself cannot be subjected to physical experimentation. *We cannot re-run the universe with the same or altered conditions to see what would happen if they were different, so we cannot carry out scientific experiments on the universe itself.* Furthermore,

Thesis A2: The universe cannot be observationally compared with other universes. *We cannot compare the universe with any similar object, nor can we test our hypotheses about it by observations determining statistical properties of a known class of physically existing universes.*

Where this all becomes of observational relevance is in the idea of *cosmic variance* [Dodelson, 2003, pp. 241, 343]. The theory of structure formation in the early universe makes statistical predictions only (it cannot attempt to predict the specific structures that will actually be formed). Testing the theory compares our universe to a theoretical ensemble of universes, and declares a variance between what is measured in the actual universe and the expected properties based on the ensemble of models. If this variance is small enough, a deviation from expected values is pronounced as a statistical deviation, i.e. of no physical significance — we do not need to explain it any further; if it is large, it needs explanation. This is a key issue for example in the analysis of the CBR anisotropy observations [White *et al.*, 1993; Kamionkowski and Loeb, 1997]. The power spectrum of the CBR as measured by WMAP is less than expected at large angular scales (Sec. 2.3.7). One school of thought claims this is just a statistical fluctuation; another that it needs explanation, and might for example be evidence of a small universe [Luminet *et al.*, 2003; Luminet, 2005]. This debate arises because there is just one universe, and on large angular scales there are just a few measurements that can possibly be made (on small angular scales we can make many measurements and so this uncertainty becomes very small).

Consequent on **A1** and **A2**,

Thesis A3: The concept of 'Laws of Physics' that apply to only one object is questionable. *We cannot scientifically establish 'laws of the universe' that might apply to the class of all such objects, for we cannot test any such proposed law except in terms of being consistent with one object (the observed universe).*

This is insufficient: one observational point cannot establish the nature of a causal relation. Indeed the concept of a 'law' becomes doubtful when there is only one given object to which it applies [Munitz, 1962]. The basic idea of a physical law is that it applies to a set of objects all of which have the same invariant underlying behaviour (as defined by that law), despite the apparent variation in properties in specific instances, this variation resulting from varying initial conditions for the systems on which the law acts. This understanding is tested by physical experiments in which initial conditions for evolution of a set of similar systems are varied, and observations by which the statistical nature of a set of objects of the same broad kind is investigated. Neither is possible in the case of cosmology.

The laws of physics apply locally to the objects in the cosmos, and determine the evolution of the cosmos as a whole when locally applied everywhere with suitable initial/boundary conditions imposed (in the case of the RW models, via the Friedmann equation for example). Apart from this, we cannot establish higher-level effective laws that apply to all universes and determine their structure, as we can at all other levels of the hierarchy of complexity. All that we can do at this level of structure is observe and analyze the one unique object that exists. This is expressed by McCrea as follows: "When we speak of the other solutions of the equations of stellar structure, besides the one we are interested in at the moment, as representing systems that could exist, we mean that they could exist in the universe as we know it. Clearly no such attitude is possible towards the universe itself" [McCrea, 1953].

Since the restriction of a global solution to a local neighborhood is also a solution, we have zillions of "mini-universe" on which to test the laws that control the local nature of the universe. But a mini-universe is not the universe itself; it is a small part of the whole. By examining these "mini-universes" and seeing if they are essentially the same everywhere, we can to some degree check firstly *that the laws of physics are the same everywhere in the universe* (a key feature of all cosmological analysis, cf. Sec. 7.1), and secondly *that the universe is spatially homogeneous* (this is discussed in depth below, see Sec. 4.2.2). But the latter feature is what has to be *explained* by a 'law of the universe'; verifying homogeneity does not explain why it is the case; this comes about because of specific initial conditions, which some suggest are due to hypothesized 'laws of the universe', applicable to the whole rather than to its parts. Finally,

Thesis A4: The concept of probability is problematic in the context of existence of only one object. *Problems arise in applying the idea of probability to cosmology as a whole — it is not clear that this makes much sense in the context of the existence of a single object which cannot be compared with any other existing object.*

But a concept of probability underlies much of modern argumentation in cosmology. Talk of 'fine tuning' for example is based on the use of probability (it is a way of saying something is improbable). This assumes both that things could have been different, and that we can assign probabilities to the set of unrealized possibilities in an invariant way. The issue here is to explain in what sense they could have been different with well-defined probabilities assigned to the different theoretical possibilities, if there is indeed only one universe with one set of initial conditions fixed somehow before physics came into being, or perhaps as physics came into being. We cannot scientifically establish laws of creation of the universe that might determine such initial conditions and resulting probabilities. If we use a Bayesian interpretation, which some suggest can be meaningfully applied to only one object [Garrett and Coles, 1993], the results depend on our 'prior knowledge', which in this case can be varied by changing our initial pre-physics assumptions. Related issues arise concerning the meaning of 'the wave function of the universe', at the heart of quantum cosmology. This wave function gives no unique prediction for any specific single universe.

Two comments on the above. First, *it is useful to distinguish between the experimental sciences — physics, chemistry, microbiology for example — on the one hand, and the historical and geographical sciences — astronomy, geology, evolutionary theory for example, on the other.* It is the former that are usually in mind in discussions of the scientific method. The understanding in these cases is that we observe and experiment on a class of identical or almost identical objects and establish their common behaviour. The problem then resides in just how identical those objects are. Quarks, protons, electrons, are all exactly identical to each other, and so have exactly the same behaviour (indeed this feature underlies well-tested quantum statistics). All DNA molecules, frogs, human beings, and ecosystems are somewhat different from each other, but are similar enough nevertheless that the same broad descriptions and laws apply to them; if this were not so, then we would be wrong in claiming they belonged to the same class of objects in the first place. Water molecules, gases, solids, liquids are in an intermediate category — almost identical, certainly describable reliably by specific physical and chemical laws.

As regards the geographical and historical sciences, here one explicitly studies objects that are unique (the Rio Grande, the continent of Antarctica, the Solar System, the Andromeda galaxy, etc.) or events that have occurred only once (the origin of the Solar System, the evolution of life on Earth, the explosion of SN1987a, etc.). Because of this uniqueness, comment **A1** above applies in these cases also: we can only observe rather than experiment; the initial conditions that led to these unique objects or events cannot be altered or experimented with. However comment **A2** does not apply: at least in principle, there is a class of similar objects out there (other rivers, continents, planetary systems, galaxies, etc.) or similar events (the origin of other galaxies, the evolution of other planetary systems, the explosion of other supernovae, etc.) which we can observe and compare with our specific exemplar, also carrying out statistical analyses on many such cases to determine underlying patterns of regularity; and in this respect these topics differ from cosmology.

If we truly cannot carry out such analyses — that is, if **A2** applies as well in some particular case — then that subject partakes in this respect of the nature of cosmology. One may claim that *the dividing line here is that if we convince ourselves that some large-scale physical phenomenon essentially occurs only once in the entire universe, then it should be regarded as part of cosmology proper*; whereas if we are convinced it occurs in many places or times, even if we cannot observationally access them (e.g. we believe that planets evolved around many stars in other galaxies), then study of that class of objects or events can be distinguished from cosmology proper precisely because there is a class of them to study. The second comment is that some workers have tried to get around this set of problems by essentially *denying the uniqueness of the universe*. This is done by proposing the physical existence of 'many universes' to which concepts of probability can be properly applied (cf. Sec. 2.7.2), envisaged either as widely separated regions of a larger universe with very different properties in each region (as in chaotic inflation for example), as multiple realizations of quantum outcomes, or as an ensemble of completely disconnected universes — there is no physical connection whatever between them — in which all possibilities are realized. We return to this in Sec. 9.2.

4 ISSUE B: THE LARGE SCALE OF THE UNIVERSE IN SPACE AND TIME.

The problems arising from the uniqueness of the universe are compounded by its vast scale in both space and time, which poses major problems for observational cosmology. We therefore need to adduce various Principles in addition to the observations, in order to attain unique models: theory comes in as basis for interpreting observations.

4.1 Observations in a large scale universe

The distance to the nearest galaxy is about 10^6 light years, that is about 10^{24}cm., while the size of the earth is about 10^9cm. The present size of the visible universe is about 10^{10} light years, that is about 10^{28}cm. This huge size relative to our own physical scale (about 10^2cm) places major constraints on our ability to observe distant regions (and certainly prevents us experimenting with them). The uniqueness of cosmology in this respect is that it deals with this scale: the largest with which we can have causal or observational contact.

Thesis B1: Astronomical observations are confined to the past null cone, fading with distance. *We can effectively only observe the universe, considered on a cosmological scale, from one space-time event. Visual observations are possible only on our past light cone, so we are inevitably looking back into the past as we observe to greater distances. Uncertainty grows with distance and time.*

The vast scale of the universe implies we can effectively only view it from one spacetime event ('here and now') [Ellis, 1971a; Ellis, 1975]. If we were to move away from this spatial position at almost the speed of light for say $10,000$ years, we would not succeed in leaving our own galaxy, much less in reaching another one; and if we were to start a long-term astronomical experiment that would store data for say $20,000$ years and then analyze it, the time at which we observe the universe would be essentially unchanged (because its age is of the order of 10^{10} years: the extra time would make a negligible difference). This is quite unlike other geographic sciences: we can travel everywhere on earth and see what is there. The situation would be quite different if the universe were much smaller. Given its actual scale, such that we are now seeing galaxies whose present distance from us is about 10^9 light years, the effect is as if we were only able to observe the earth from the top of one mountain, and had to deduce its nature from those observations alone [Ellis, 1975].

Because we can only observe by means of particles — photons, massless neutrinos, gravitons — travelling to us at the speed of light, astronomical observations of distant sources and background radiation by telescopes operating at all wavelengths (optical, infrared, ultraviolet, radio, X-ray) are constrained to rays lying in our past light cone. These allow detailed observations (including visual pictures, spectral information, and polarization measurements) of matter as it intersects our past light cone. In observing distant regions, we can also aspire to use neutrino and gravitational wave telescopes, and perhaps cosmic rays, also representing information coming to us at the speed of light or less. However all our detailed data about distant regions is gathered along our past light cone.

As a consequence, three interrelated problems occur in interpreting the astronomical observations. The first is that (because we can only view the universe from one point) *we only obtain a 2-dimensional projection on the sky of the 3-dimensional distribution of matter in the universe.* To reconstruct the real distribution, we need reliable distance measurements to the objects we see. However because of variation in the properties of sources, most are not reliable standard candles or standard size objects to use in calibrating distances, and in these cases we have to study statistical properties of classes of sources to estimate distances.

Second, *we necessarily see distant galaxies and other objects at earlier times in their history* (where their world lines intersect this past light cone).[21] Thus cosmology is both a geographic and a historical science combined into one: we see distant sources at an earlier epoch, when their properties may have been different. As we are looking back in the past, source evolution must be taken into account; their properties at the time they emitted the light may be quite different from their properties now. We can only determine the distances of objects if we understand this evolution; but in practice it is one of the unknowns we have to try to determine (cf. Sec. 4.2.3).

Third, distant sources appear very small and very faint, both because of their physical distance, and because their light is highly redshifted (due to the expansion of the universe). Simply detecting them, let alone determining their characteristics, becomes rapidly more difficult with distance. Furthermore absorption by intervening matter can interfere with light from distant objects. The further back we look, the worse these problems become; thus our reliable knowledge of the universe decreases rapidly with distance [Ellis, 1975].

The situation is however improved by the availability of geological-type data [Hoyle, 1960]; that is, the present-day status of rocks, planets, star clusters, galaxies, and so on, which contains much information on the past history of the matter comprising those objects. Thus we can obtain detailed information on conditions near our past world-line in spacetime [Ellis, 1971a; Ellis, 1975] at very early times if we can interpret this data reliably, for example by relating theories of structure formation to statistical studies of source properties.

Thesis B2: 'Geological' type observations can probe the distant past of our past world line. *Physical and astrophysical observations tell us about conditions near matter world-lines in the far distant past. They can be used also to investigate the far distant past of more distant objects.*

This involves us in physical cosmology: namely the study of the evolution of structures in the universe, tested by comparison with astronomical observation. Particularly useful are measurements of the abundances of elements which resulted

[21] For example we see the Andromeda galaxy as it was two million years ago, long before humans existed on Earth [Silk, 2005].

from nucleosynthesis in the Hot Big Bang, giving us data about conditions long before decoupling (Sec. 2.2.2). If we can obtain adequate quality data of this kind for objects at high redshifts, we can use this to probe conditions very early on in their histories at some distance from our past worldline. Encouraging in this regard is the possibility of determination of element abundances at high redshift [Dodelson, 2003, pp. 11-12]; [Pettini, 1999]).

4.2 Determining Spacetime Geometry: Observational Limits.

The unique core business of observational cosmology is determining the large-scale geometry of everything there is, or at least of everything we can observe.

4.2.1 Direct determination versus theory based approaches

One can go about this in a direct manner: trying to determine the geometry of the universe directly from observations (assuming one has some understanding of the sources observed). The way this can be done (curiously known as the 'inverse approach') has been fully characterized [Kristian and Sachs, 1966; Ellis *et al.*, 1985]; indeed there is an interesting result here, namely

Observational Cosmology Theorem: *The data in principle available on our past null cone from astronomical observations is just necessary and sufficient to determine the space-time geometry on that null cone* [Ellis *et al.*, 1985]. *From this data one can in principle determine the space time in the past of the null cone and, if a no-interference conditions is assumed, to its future.*

However this is difficult to carry out both because of the problem of estimating distances for all observed sources, requiring a knowledge of the nature of the sources (Sec. 4.2.3),[22] and because of the serious difficulty in obtaining some of the needed data (which include apparent distortions of all distant objects, and the transverse velocities of all observed matter). The further we observe down the past light cone, the larger the uncertainty becomes. This direct observational approach, where no prior model is assumed for the space-time geometry, has been pursued to some degree (and in essence underlies for example the observational studies that discovered large-scale structure such as the great walls and voids). Nevertheless it is not widely adopted as an overall approach to cosmology, both because of these observational difficulties, but also because it has little explanatory value; it just tells us what the geometry and matter distribution is, but not why it is of that nature.

The usual option in cosmology proper is rather to use a theory-based approach: we *a priori* assume a model based on a space-time geometry with high symmetry (usually a FL model, see Sec. 2.1), and then determine its essential free parameters from comparison of theoretical relations with astronomical observations (Sec. 2.3.3). Detailed observations of the matter distribution and large-scale velocities

[22] The link between observations and models always requires some theory, and is never direct.

as well as CBR anisotropies then help us determine deviations from the exact model, both statistically (an astrophysical description [Dodelson, 2003]) and in detail (an astronomical description [Ellis and Stoeger, 1987]).

4.2.2 Indirect determination: justifying a Friedmann-Lemaître geometry

The standard models of cosmology are the Friedmann-Lemaître (FL) family of universe models that are exactly spatially homogeneous and isotropic everywhere (Sec. 2.1). They are easy to understand, and have tremendous explanatory power; furthermore their major physical predictions (the existence of blackbody CBR and specific light element production in the early universe) seem confirmed. The issue is, to what degree does observational data uniquely indicate these universe models for the expanding universe geometry? Here one is assuming a large enough averaging scale for spatial homogeneity to be valid; this scale should be explicitly indicated [Ellis, 1984] (it is about 100 Mpc at present [Dodelson, 2003]).[23] These are the background models for cosmology; perturbed FL models then characterize the nature of deviations from the exact FL geometry that are expected on smaller scales (Sec. 2.5.2).

The key feature here is the observed *isotropy about our location* (Sec. 2.3.1). Considered on a large enough angular scale, astronomical observations are very nearly isotropic about us, both as regards source observations and background radiation; indeed the latter is spectacularly isotropic, better than one part in 10^4 after a dipole anisotropy, understood as resulting from our motion relative to the rest frame of the universe, has been removed [Partridge, 1995]. Because this applies to all observations (in particular, there are not major observed matter concentrations in some other universe region), this establishes that in the observable region of the universe, to high accuracy *both the space-time structure and the matter distribution are isotropic about us.* We can easily construct exact spherically symmetric universe models [Bondi, 1947; Ellis and van Elst, 1999a], as indicated by these observations. In general they will be spatially inhomogeneous, with our Galaxy located at or near the centre; this is currently a philosophically unpopular proposal, but is certainly possible. The question is whether we can give convincing observational evidence for spatial homogeneity in addition to the spherical symmetry. Various arguments are used for this purpose.

(a) *The cosmological principle* [Bondi, 1960; Weinberg, 1972]: Just assume spatial homogeneity because it is the simplest case and you don't need anything more complex on the basis of current data. We simply adopt a philosophical principle as the basis of argument. This is essentially an *a priori prescription for initial conditions for the universe* (a universe that initially has a RW geometry will have that geometry at later times, because symmetries of

[23]There exist *hierarchical models* where neither the fluid approximation nor homogeneity is ever attained at any scale because of their fractal nature [de Vaucouleurs, 1970]. The regularity of the observed galactic motions, as evidenced by the (m, z) relations, speaks against these models, as do large-scale observations of the matter distribution [Peebles, 1993a].

the initial data are preserved by the Einstein equations [Hawking and Ellis, 1973]); but it is not usually expressed that way.

(b) *FL observational relations*: If we could show that the source observational relations had the unique FL form (26, 28) as a function of distance, this would establish spatial homogeneity in addition to the isotropy, and hence a RW geometry [Ellis *et al.*, 1985]. This is essentially what is done for example in using number counts towards establishing spatial homogeneity [Hubble, 1936]. However because of Thesis **B1** above, the observational problems mentioned earlier — specifically, unknown source evolution — prevent us from carrying this through: we cannot measure distances reliably enough. Astrophysical cosmology could resolve this in principle, but is unable to do so in practice. Indeed the actual situation is the inverse: *taking radio-source number-count data at its face value, without allowing for source evolution, contradicts a RW geometry.*

In the face of this, the usual procedure is to assume that spatial homogeneity is known in some other way, and deduce the source evolution required to make the observations compatible with this geometric assumption [Ellis, 1975]. It is always possible to find a source evolution that will achieve this [Mustapha *et al.*, 1998]. Thus attempts to observationally prove spatial homogeneity this way fail; indeed an alternative interpretation would be that this data is evidence of spatial inhomogeneity, i.e. that we live in a spherically symmetric inhomogeneous universe where we are situated somewhere near the centre, with the cosmological redshift being partly gravitational, cf. [Ellis *et al.*, 1978] (and conceivably with a contribution to the CBR dipole from this inhomogeneity if we are a bit off-centre). Similarly the supernova data usually understood as implying the existence of a cosmological constant (Sec. 2.3.5) could also be interpreted in this way as evidence of inhomogeneity, without the need for 'dark energy'. Most people regard such proposals as very unappealing — but that does not prove they are incorrect.

(c) *Physical arguments*: One can claim that physical processes such as inflation (Sec. 2.6) make the existence of almost-RW regions highly likely, indeed much more probable than spherically symmetric inhomogeneous regions. This is a viable argument, but we must be clear what is happening here — we are replacing an observational test by a theoretical argument based on a physical process that may or may not have happened (for there is no definitive observational proof that inflation indeed took place). It is strongly bolstered because predictions for the detailed pattern of CBR anisotropy on small scales [Hu and Sugiyama, 1995b], based on the inflationary universe theory, have been confirmed [Perkins, 2005]; but that argument will only become rigorous if it is shown that spherically symmetric inhomogeneous models (with or without inflation) cannot produce similar patterns of anisotropy. But they probably can, because the acoustic oscillations that lead to the characteristic predicted anisotropy patterns in fact take place after inflation, and can equally happen if suitable initial conditions occur without a previous inflationary phase.

What about alternative observational routes? Another proposal is,

(d) *Uniform thermal histories*: the idea is to use the uniformity in the nature of the objects we see in the sky (we see the same types of galaxy at large distances, for example) to deduce they must have all undergone essentially the same thermal history, and then to prove from this homogeneity of thermal histories that the universe must be spatially homogeneous. For example, observations showing that element abundances at high redshift in many directions are the same as locally, are very useful in constraining inhomogeneity by showing that conditions in the very early universe at the time of nucleosynthesis must have been the same at distant locations in these directions [82]. However turning this idea into a proper test of homogeneity has not succeeded so far: indeed it is not clear if this can be done, because some (rather special) counter-examples to this conjecture have been found [Bonnor and Ellis, 1986]. Nevertheless the approach could be used to give evidence against spatial homogeneity: for example, if element abundances were measured to be different at high redshifts in any direction [Pettini, 1999; Sigurdson and Furlanetto, 2005], or if ages of distant objects were incompatible with local age estimates [Jain and Dev, 2005].

Finally the argument for spatial homogenity that is most generally accepted:

(e) *Isotropy everywhere:* If all observers see an isotropic universe, then spatial homogeneity follows [Walker, 1944; Ehlers, 1993; Ellis, 1971a]; indeed homogeneity follows if only three spatially separated observers see isotropy. Now we cannot observe the universe from any other point, so we cannot observationally establish that far distant observers see an isotropic universe. Hence the standard argument is to assume a *Copernican Principle:* that we are not privileged observers. This is plausible in that all observable regions of the universe look alike: we see no major changes in conditions anywhere we look. Combined with the isotropy we see about ourselves, this implies that *all observers see an isotropic universe*, and this establishes a RW geometry [Walker, 1944; Ellis, 1971a; Hawking and Ellis, 1973]. This result holds if we assume isotropy of *all* observations; a powerful enhancement was proved by Ehlers, Geren, and Sachs [Ehlers *et al.*, 1968; Hawking and Ellis, 1973], who showed that if one simply assumes isotropy of freely-propagating radiation about each observer in an expanding universe domain,[24] the result follows from the Einstein and Liouville equations; that is,

[24]This result does not hold in a static universe, for then the radiation temperature depends only on the potential difference between the emitter and observer, hence the radiation is isotropic everywhere even if the universe inhomogeneous, cf. [Ellis *et al.*, 1978].

> **EGS Theorem:** *Exact isotropy of the CBR for every geodesically moving fundamental observer at each point in an expanding universe domain U implies an exact RW geometry in U.*

Thus we may establish spatial homogeneity by assuming a weak Copernican principle: we are not in a privileged position where the CBR just happens to be highly isotropic by chance; hence all comoving observers may be assumed to measure highly isotropic CBR, and the result follows. This is currently the most persuasive observationally-based argument we have for spatial homogeneity.

A problem is that it is an exact result, assuming exact isotropy of the CBR; is the result stable? Indeed it is: *almost-isotropy of freely-propagating CBR for an expanding family of geodesically-moving fundamental observers everywhere in some region proves the universe geometry is almost-RW in that region* [Stoeger *et al.*, 1995]. Thus the result applies to the real universe — provided we make the Copernican assumption that all other observers, like us, see almost isotropic CBR. And that is the best we can do at present. Weak tests of the isotropy of the CBR at other spacetime points come from the Sunyaev-Zel'dovich effect [Goodman, 1995] and from CBR polarization measurements [Kamionkowski and Loeb, 1997], giving broad support to this line of argument but not enough to give good limits on spatial inhomogeneity.

The observational situation is clear:

Thesis B3: Establishing a Robertson-Walker geometry for the universe relies on plausible philosophical assumptions. *The deduction of spatial homogeneity follows not directly from astronomical data, but because we add to the observations a philosophical principle that is plausible but untestable.*

The purpose of the above analysis is not to seriously support the view that the universe is spherically symmetric and inhomogeneous, as is allowed by the observations, but rather to show clearly the nature of the best observationally-based argument by which we can (quite reasonably) justify the assumption of spatial homogeneity.

Accepting this argument, the further question is, *in which spacetime regions does it establish a RW-like geometry*? The CBR we detect probes the state of the universe from the time of decoupling of matter and radiation (at a redshift of about 1100) to the present day, within the visual horizon. The argument from CBR isotropy can legitimately be applied for that epoch. However, it does not necessarily imply isotropy of the universe at much earlier or much later times, because there are spatially homogeneous anisotropic perturbation modes that are unstable in both directions of time; and they will occur in a generic situation. Indeed, if one examines the Bianchi (spatially homogeneous but anisotropic) universes, using the powerful tools of dynamical systems theory, one can show that *intermediate isotropisation* can occur [Wainwright and Ellis, 1996; Wainright *et al.*, 1998]: *despite being highly anisotropic at very early and very late times, such models can mimic a RW geometry arbitrarily closely for an arbitrarily long time*, and hence can reproduce within the errors any set of FL-like observations. We can obtain strong limits on the present-day strengths of these anisotropic modes from CBR anisotropy measurements and from data on element abundances, the latter being a powerful probe because (being of the 'geological' kind) they can test conditions at the time of element formation, long before decoupling. But however low these observational limits, anisotropic modes can dominate at even earlier times as well as at late times (long after the present). If inflation took place, this conclusion is reinforced: it washes out any information about very early universe anisotropies and inhomogeneities in a very efficient way.

As well as this time limitation on when we can regard homogeneity as established, there are major spatial limitations. The above argument does not apply far outside the visual horizon, for we have no reason to believe the CBR is highly isotropic there. Indeed if chaotic inflation is correct, conditions there are not the same.

4.2.3 Determining the RW parameters

Given that a RW geometry is a good description of the observable universe on a large scale, the further issue is what are the best-fit parameters that characterize it, selecting the specific universe we observe from the family of all FL models (Sec. 2.1). Important observational issues are:

- Determining the Hubble parameter H_0, which sets the overall scale of the observed universe region.

- Determining the trio of the density parameter Ω_0, deceleration parameter q_0, and cosmological constant Λ (or equivalently the density parameter Ω_Λ), which are the major defining characteristics of a specific FL model. The CBR data, supernova observations, deep number counts, source covariance functions, velocity measurements, and gravitational lensing observations can determine these quantities.

- Determining the sign of the curvature k, showing whether the universe has closed spatial sections and also whether it is possible for it to recollapse in the future or not. Analyses of the observations should always attempt to determine this sign, and not assume that $k = 0$ (as is often done) [Wright, 2006].

- Various parameters are used to characterize the nature of dark matter (Sec. 2.3.6) and dark energy (Sec. 2.3.5). As their dynamics is unknown, these too have to be determined observationally.

We only obtain good estimates of these quantities by the observational relationships characterized above (Sec. 2.3.3) using statistical analysis of the classes of objects we observe. Problems arise because of our lack of adequate theories of their historical development.

Thesis B4: Interpreting cosmological observations depends on astrophysical understanding. *Observational analysis depends on assessing a variety of auxiliary functions characterizing the sources observed and the observations made. These introduce further parameters that have to be observationally or theoretically determined, allowing considerable freedom in fitting specific models to the observations. Physical cosmology aims to characterize perturbed FL models (which account for structure formation) rather than just the background exactly smooth FL models; this introduces further parameters to be determined.*

It is useful here to distinguish between methods aimed at determining the properties of the background (zeroth order) FL model directly, and those aimed at determining properties of the perturbations of these models [Tegmark, 2002]. Methods for determining the parameters of the background model (Sec. 2.1) depend on assuming properties of the distance indicators used (galaxies, radio sources, etc.). They will have their own properties (brightness profiles, luminosities, physical sizes, spectra, etc.) and dynamical evolution; but these are often not well understood, and will have to be represented in a parametric way (e.g. by parameters describing luminosity evolution). In each case we end up assuming important aspects of the astrophysics and evolutionary histories of the objects observed, which are not part of the cosmological model proper. The statistical properties of the sources observed are also characterized by parametrized functions (e.g. the luminosity function characterizing the numbers of galaxies in each luminosity class) that have to be known in order to analyze the observations. This situation is an exemple of Lakatos' view of how scientific programmes work, with a belt of auxiliary hypotheses interposing between the core theoretical proposal and the data used to test it [Lakatos, 1980]. This makes the analysis rather model-dependent, where the models are only indirectly related to the background model — their explanation is the aim of astrophysics rather than cosmology. Thus if observational results disagree with a particular cosmological model, one can always claim it is the understanding of the auxiliary hypotheses that is at fault rather than the model being proposed [Lakatos, 1980].

By contrast, many of the methods of estimating Ω_0 (and to some degree Λ) depend on studying the growth and nature of inhomogeneities in the universe, that is they investigate perturbed FL models (Sec. 2.5.2), whose properties of course depend on the background model, but introduce a whole set of further functions and parameters describing the perturbations [Dodelson, 2003], for example the angular correlation function for matter (or its Fourier transform, the 2-dimensional power spectrum), the power spectrum of density fluctuations [Tegmark, 2002], red-shift space correlation functions [Peebles, 1993a; Eisenstein *et al.*, 2005a], and correlation function for velocities [Dodelson, 2003]. Associated parameters include a scalar *spectral index* (characterizing the spectrum of physical sizes of inhomogeneities), the *bias parameter* b (expressing how galaxy formation is biassed towards density peaks in the inhomogeneities [Dodelson, 2003, p. 280]) and the *initial fluctuation magnitudes* Q (the seeds for structure formation). Determining these parameters is part of the task of cosmology proper: to fully characterize the perturbed cosmological model, *we aim to determine both the background parameters and the quantities describing the perturbations.* Model selection then depends on the parameters used to describe them — what is assumed known, and what is to be determined [Liddle, 2004; Scott, 2005]. For example, standard inflationary theory predicts a scale-invariant spectrum of Gaussian perturbations; do we test that assumption, or take it for granted? This comes up in the issue of what 'priors' are assumed when conducting statistical tests.

4.2.4 Consistency tests

A key question for cosmology is *what kinds of observations provide critical tests of the standard FL models.* If there were no observations that could disprove them, the subject would be of questionable scientific status. An important such test is obtaining *estimates of the age of the universe* t_0, which is dependent on H_0, Ω_0, and Λ, and comparing them with estimates of the ages of objects in the universe (determined on astrophysical grounds):

Thesis B5: A crucial observational test for cosmology is that the age of the universe must be greater than the ages of stars. *The tension between the age of the universe and ages of stars is one area where the standard models are vulnerable to being shown to be inconsistent, hence the vital need to establish reliable distance scales, basic to estimates of both* H_0 *and the ages of stars, and good limits on* Λ. *Other consistency tests help confirm the standard model and consolidate cosmology's standing as an empirical science.*

At present this age issue is acceptable for local objects, because of a recent revision of our distance scale estimates [Harris *et al.*, 1998], assisted by data that Λ is positive [Perlmutter *et al.*, 1998]; but continued vigilance is needed on this front, particularly as there are indications of problems for high redshift objects [Jain and Dev, 2005]. If this ever became serious we might have to resort to spherically symmetric inhomogeneous models rather than spatially homogeneous models, with the 'bang time' (characterizing the start of the universe) dependent on distance from us [Mustapha *et al.*, 1998].

Note that this issue is crucially unlike the case of the large angle CBR anisotropies (Sec. 2.3.7): *the low CBR anisotropies at large angular scales can as a last resort be dismissed as a statistical fluke; the age issue cannot.* It is to do with the internal consistency of individual cosmological models, not with probabilities. Thus it is a plus for cosmology that the age issue exists. Other consistency tests include

- Showing that the CBR temperature T_{cbr} varies with redshift according to $T_{cbr} = 2.75\,(1+z)$ [Meyer, 1994];
- Confirming that helium abundances are consistent with a primordial value of 25% at large distances (high redshifts) in all directions [Dodelson, 2003, pp. 11-12]; also [Pettini, 1999; Sigurdson and Furlanetto, 2005]; and

- Checking that there is a 2% number count dipole parallel to the CBR dipole for all cosmological sources [Ellis and Baldwin, 1984].

4.3 The hidden universe

If we do not live in a small universe (Sec. 4.3.1), the further essential point is that the region of the universe we can observe is restricted, firstly because we cannot see to earlier times than the LSS (the universe was opaque before then (see Sec. 2.2)), and secondly because a finite time has elapsed since the universe became transparent to radiation, and light can only have travelled a finite distance in that time. As no signal can travel to us faster than light, we cannot receive any information from galaxies more distant than our visual horizon [Ellis and Stoeger, 1988]. The most distant matter we can observe is that which emitted the CBR (Sec. 2.4.2).

Thesis B6: Observational horizons limit our ability to observationally determine the very large scale geometry of the universe. *We can only see back to the time of decoupling of matter and radiation, and so have no direct information about earlier times; and unless we live in a 'small universe', most of the matter in the universe is hidden behind the visual horizon. Conjectures as to its geometry on larger scales cannot be observationally tested. The situation is completely different in the small universe case: then we can see everything there is in the universe, including our own galaxy at earlier times.*

The key point here is that unless we live in a small universe, *the universe itself is much bigger than the observable universe.* There are many galaxies — perhaps an infinite number — at a greater distance than the horizon, that we cannot observe by any electromagnetic radiation. Furthermore no causal influence can reach us from matter more distant than our particle horizon — the distance light can have travelled since the creation of the universe, so this is the furthest matter with which we can have had any causal connection [Rindler, 1956; Hawking and Ellis, 1973; Tipler *et al.*, 1980]. We can hope to obtain information on matter lying between the visual horizon and the particle horizon by neutrino or gravitational radiation observatories; but we can obtain no reliable information whatever about what lies beyond the particle horizon. We can in principle feel the gravitational effect of matter beyond the horizon because of the force it exerts (for example, matter beyond the horizon may influence velocities of matter within the horizon, even though we cannot see it). This is possible because of the constraint equations of general relativity theory, which are in effect instantaneous equations valid on spacelike surfaces.[25] However we cannot uniquely decode that signal to determine what matter distribution outside the horizon caused it: a particular velocity field might be caused by a relatively small mass near the horizon, or a much larger

[25] They are valid at any late time in a solution of the EFE because they were valid initially — the initial data must satisfy constraint equations — and once they are satisfied, the constraints are preserved by the dynamic field equations.

mass much further away [Ellis and Sciama, 1972]. Claims about what conditions are like on very large scales — that is, much bigger than the Hubble scale — are unverifiable [Ellis, 1975], for we have no observational evidence as to what conditions are like far beyond the visual horizon. The situation is like that of an ant surveying the world from the top of a sand dune in the Sahara desert. Her world model will be a world composed only of sand dunes — despite the existence of cities, oceans, forests, tundra, mountains, and so on beyond her horizon.

It is commonly stated that if we live in a low-density universe and the cosmological constant vanishes, the universe has infinite spatial sections. However this deduction only applies if firstly the RW-like nature of the universe within the past light cone continues to be true indefinitely far outside it, and secondly the space sections have their 'natural' simply-connected topology — and there is no way we can obtain observational evidence that these conditions are both true. In contrast to this, in chaotic inflationary models (Sec. 2.6), it is a definite prediction that the universe will not be like a RW geometry on a very large scale — rather it will consist of many RW-like domains, each with different parameter values, separated from each other by highly inhomogeneous regions outside our visual horizon [Linde, 1990], the whole forming a fractal-like structure. This prediction is just as untestable as the previously prevalent assumption (based on a Cosmological Principle) that the universe is RW-like on such scales [Bondi, 1960; Weinberg, 1972]. Neither can be observationally confirmed or denied. The same issue arises in an even more extreme form in relation to the idea of a multiverse. We return to this below, see Sec. 9.2.

4.3.1 Small universes

There is one case where this kind of spatial observational limit does not obtain. This is when a *Small Universe* occurs, that is, a universe which closes up on itself spatially for topological reasons [Ellis, 1971b], and does so on such a small scale that we have seen right round the universe since the time of decoupling. Then we can see all the matter that exists, with multiple images of many objects occurring [Ellis and Schreiber, 1986]. This possibility is observationally testable by examining source statistics, by observation of low power in the large angle CBR anisotropies, and by detecting identical temperature variation on various circles in the CBR sky [Lachièze *et al.*, 1995]. There are weak hints in the observed CBR anisotropies (the lack of power on large angular scales) that this could actually be the case [Luminet *et al.*, 2003; Luminet, 2005], but this is not solidly confirmed. Checking if the universe is a small universe or not is an important task; the nature of our observational relationship to the universe is fundamentally different if it is true [Ellis and Schreiber, 1986].

4.4 The observed universe

The observable part of the universe (i.e. back to the visual horizon) is strictly limited, and we have already seen most of it. We can only observe distant objects by electromagnetic radiation at all wavelengths, by neutrinos, and by gravitational waves. We already have very complete broad coverage of the entire sky by electromagnetic observations at all wavelengths right back to the surface of last scattering, which is the limit of what will ever be observable by electromagnetic radiation. Detailed observations (such as the Hubble Deep Field) are available for restricted domains in angle and depth. Detailed observations at suitable wavelengths are beginning to discern what lies behind the Milky Way, which tends to obscure a substantial fraction of the sky. It is unlikely there are many new astronomical phenomena undiscovered in this observable region, although it will be crucial determining more detailed features of the phenomena we have already discovered (e.g. the nature of dark matter and dark energy).

Thesis B7: We have made great progress towards observational completeness. *We have already seen most of the part of the universe that is observable by electromagnetic radiation. It is plausible that not many new astronomical phenomena remain to be discovered by us observationally; we will determine more details (so understanding more about what we have seen) and see more objects, but not discover many new kinds of things.*

Indeed Harwit [1984] has used the multiplicity of discovery of specific astronomical phenomena to estimate how many new essentially different such phenomena there are still waiting to be discovered.

Neutrinos and gravitational waves will in principle allow us to peer back to much earlier times (the time of neutrino decoupling and the quantum gravity era respectively), but are much harder to observe at all, let alone in useful directional detail. Nevertheless the latter has the potential to open up to us access to eras quite unobservable in any other way. Maybe they will give us unexpected information on processes in the very early universe which would count as new features of physical cosmology.

5 ISSUE C: THE UNBOUND ENERGIES IN THE EARLY UNIVERSE

The analogous problems for physical cosmology arise because energies occurring in the Hot Big Bang early universe phase (Sec. 2.2) are essentially unbounded, so the highest energies we can attain in particle accelerators cannot reach the levels relevant to very early times. The uniqueness of cosmology in this regard is that it is the only science contemplating spacetime regions that have experienced such high energies, and with which we are in intimate causal contact despite the huge timescales involved — indeed events at those early times determined much of what we see around us today. The nuclear reactions underlying nucleosynthesis are well understood, and their cross-sections reasonably well-known; the processes of baryosynthesis and quark-gluon recombination are reasonably understood and are on the border of being testable; but physical processes relevant at earlier times are inaccessible to testing by laboratory or accelerator-based experiment. The *Physics Horizon* by definition separates those aspects of physics we can hope to test by high-energy experiments on Earth or in the Solar System, from those where it is reasonable to expect no such test will ever be possible:

Thesis C1: The Physics Horizon limits our knowledge of physics relevant to the very early universe. *We cannot experimentally test much of the physics that is important in the very early universe because we cannot attain the required energies in accelerators on Earth. We have to extrapolate from known physics to the unknown and then test the implications; to do this, we assume some specific features of known lower energy physics are the true key to how things are at higher energies. We cannot experimentally test if we have got it right.*

Note that this is independent of the issue of setting of initial conditions for the universe, considered below, see Sec. 6.2: the problem arises after the initial conditions have been set and the universe is running according to invariable physical laws. We cannot be confident of the validity of the physics we presuppose then. Rather than using known physics to predict the evolution of the universe, *we end up testing proposals for this physics by exploring their implications in the early universe*, which is the only 'laboratory' where we can test some of our ideas regarding fundamental physics at the highest energies [Yoshimura, 1988]; this is particularly true in the case of quantum gravity proposals. The problem is we cannot simultaneously do this and also carry out the aim of physical cosmology, namely predicting the evolution of the early universe from known physical theory.

Our understanding of physics at those times has of necessity to be based on extrapolation of known physics way beyond the circumstances in which it can be tested. The trick is to identify which features are the key to use in that extrapolation: for example, variational principles, broken symmetries and phase changes, duality invariance, entropy limits are candidates. If we confirm our guesses for the relevant physics by their satisfactory implications for the early universe, tested in some suitable way, then this is impressive progress; but if this is the *only* way we can test the proposed physics, the situation is problematic. If the hypothesis solves only the specific issues it was designed to solve in the early universe and nothing else, then in fact it has little explanatory power, rather it is just an alternative (perhaps theoretically preferable) description of the known situation. One obtains positive observational support for a particular proposal for the relevant physics only if it predicts multiple confirmed outcomes (rather than just one), for example predicting particles that are then confirmed to exist in a laboratory, so that a single hypothesis simultaneously solves several different observational issues. Some of the options may be preferred to others on various theoretical grounds; but one must distinguish this from their having observational support. They lack physical power if they have no other testable consequences. A particular example is the inflationary universe proposal (Sec. 2.6): the supposed inflaton field underlying an inflationary era of rapid expansion in the early universe [Guth, 1981; Gibbons *et al.*, 1983; Kolb and Turner, 1990; Guth, 1997] has not been identified, much less shown to exist by any laboratory experiment. Because this field ϕ is unknown, one can assign it an arbitrary potential $V(\phi)$, this arbitrariness reflecting our inability to experimentally determine the relevant behaviour. It can be shown that virtually any desired scale evolution $S(t)$ of the universe can be attained by suitable choice of this potential [Ellis and Madsen, 1991]; and also almost any desired perturbation spectrum can be obtained by a (possibly different) suitable choice [Lidsey *et al.*, 1997]. Indeed in each case one can run the mathematics backwards to determine the required potential $V(\phi)$ from the desired outcome (Sec. 9.3.1 below). The mathematical existence of such a theoretical potential of the desired form for cosmological purposes does not by itself prove a particle or field exists with that effective potential.

Thesis C2: The unknown nature of the inflaton means that inflationary universe proposals are incomplete. *The promise of inflationary theory in terms of relating cosmology to particle physics has not been realized. This will only be the case when the nature of the inflaton has been pinned down to a specific field that experiment confirms or particle physics requires to exist.*

The very impressive achievement of inflation is that the predicted CBR anisotropy spectrum is verified and agrees with the matter power spectrum [Eisenstein *et al.*, 2005a]; but that prediction depends only on the physics from the era of tight coupling of matter and radiation to the present day, given a suitable initial fluctuation spectrum in the early universe, rather than on the specific hypothesis of an inflationary origin for that spectrum. The true clincher would be if properties of an inflationary field were predicted from the cosmology side and then confirmed in the laboratory; indeed that would count as one of the great feats of theoretical physics. This may not happen however because of the experimental problems focused on here, arising because we cannot reproduce on Earth all the conditions relevant to very early cosmology.

One key application where this issue becomes significant is in respect of the chaotic inflation theory (Sec. 2.6). As remarked above, see Sec. 4.3, its geometric predictions are observationally unverifiable. It would nevertheless be a good physical prediction if it was a more or less inevitable outcome of known and tested underlying physics. However this is not the case: the proposed underlying physics is not experimentally tested, indeed it is not even uniquely defined or associated with any specific known physical particle or field. The claim that it inevitably follows from string theory [Susskind, 2005] suffers from the problem that string theory is not a well-defined or tested part of physics.

6 ISSUE D: EXPLAINING THE UNIVERSE — THE QUESTION OF ORIGINS.

This is the unique core business of physical cosmology: explaining both why the universe has come into existence and evolved to the present very high-symmetry FL geometry on large scales, and how structures come into existence on smaller scales.

6.1 Start to the universe

Did a start to the universe happen? If so, what was its nature? This has been discussed above (Sec. 2.7.2), and the issue is unresolved. The major related question is whether the process of expansion only happens once in the life of the Universe, or occurs repeatedly. The first option is the standard model, where the entire evolution of the Universe is a once-off affair, with all the objects we see, and indeed the Universe itself, being transient objects that will burn out like dead fireworks after a firework display. In this case everything that ever happens occurs during one expansion phase of the Universe (possibly followed by one collapse phase, which could occur if $k = +1$ and the present 'dark energy' field dies away in the future). This evolution might have a singular start at a space-time singularity; a beginning where the nature of time changes character; a non-singular bounce from a single previous collapse phase; or a start from a non-singular static initial state [Mulryne *et al.*, 2005]. An alternative is that many such phases have occurred in the past, and many more will occur in the future; the Universe is a *Phoenix Universe* [Dicke and Peebles, 1979], new expansion phases repeatedly arising from the ashes of the old. While the idea of one or more bounces is an old one [Tolman, 1934], actual mechanisms that might allow this bounce behaviour have not yet been elucidated in a fully satisfactory way. A variant is the chaotic inflation idea of new expanding universe regions arising from vacuum fluctuations in old expanding regions, leading to a universe that has a fractal-like structure at the largest scales, with many expanding regions with different properties emerging out of each other in a universe that lasts forever (Sec. 2.6).

As discussed above, see Sec. 2.7.1, it is possible (if the universe has positive spatial curvature) that the quantum gravity domain can be avoided and there was no start to the universe; however this probably requires special initial conditions [Ellis and Maartens, 2004]. If a quantum gravity domain indeed occurred, we cannot come to a definite conclusion about whether there was a creation event or not because we do not know the nature of quantum gravity, nor how to reliably apply it in the cosmological context where the issue of initial conditions arises. Loop quantum gravity suggests the universe may be singularity-free [Bojowald, 2001], with bounces or a non-singular start, but that theory is unconfirmed. Tested physics cannot give a decisive answer; it is possible that *testable* physics also cannot do so.

Thesis D1: An initial singularity may or may not have occurred. *A start to the universe may have occurred a finite time ago, but a variety of alternatives are conceivable: eternal universes, or universes where time as we know it came into existence in one or another way. We do not know which actually happened, although quantum gravity ideas suggest a singularity might be avoided.*

This is a key issue in terms of the nature of the universe: a space-time singularity is a dramatic affair, where the universe (space, time, matter) has a beginning and all of physics breaks down and so the ability to understand what happens on a scientific basis comes to an end. However eternal existence is also problematic, leading for instance to the idea of Poincaré's eternal return: everything that ever happened will recur an infinite number of times in the future and has already occurred an infinite number of times in the past [Barrow and Tipler, 1984]. This is typical of the problems associated with the idea of infinity (discussed further below, see Sec. 9.3.2). *It is not clear in the end which is philosophically preferable: a singularity or eternal existence.* That decision will depend on what criteria of desirability one uses (such criteria are discussed below, see Sec. 8.1).

6.2 The issue of initial conditions

While occurrence of an initial singularity is striking in that it is a start to physics and spacetime as well as matter, whether it occurred or not is in a sense irrelevant to the key issue of what determined the nature of the universe:

Thesis D2: Testable physics cannot explain the initial state and hence specific nature of the universe. *A choice between different contingent possibilities has somehow occurred; the fundamental issue is what underlies this choice. Why does the universe have one specific form rather than another, when other forms consistent with physical laws seem perfectly possible? The reasons underlying the choice between different contingent possibilities for the universe (why one occurred rather than another) cannot be explored scientifically. It is an issue to be examined through philosophy or metaphysics.*

Even if a literal creation does not take place, as is the case in various of the present proposals, this does not resolve the underlying issue of what determined why the universe is the way it is, given that it could presumably have been otherwise. If the proposal is evolution from a previous eternal state — Minkowski space for example — then why did that come into existence, and why did the universe expansion as a bubble from that vacuum start when it did, rather than at some previous time in the pre-existent eternity? Whenever it started, it could have started before! Some attempts involve avoiding a true beginning by going back to some form of eternal or cyclic initial state, for example Tolman's series of expansion and collapse cycles [Tolman, 1934], proposals for creation of the universe as a bubble formed in a flat space-time [Tryon, 1973], Linde's eternal chaotic inflation [Linde, 1990], Veneziano's re-expansion from a previous collapse phase [Ghosh *et al.*, 1998], the ekpyrotic universe proposal [Khoury *et al.*, 2001], and theories involving foundational limits on information through a "holographic principle" [Susskind and Lindesay, 2004]. These do not avoid the ultimate problem; it can be claimed they simply postpone facing it, for one now has to ask all the same questions of origins and uniqueness about the supposed prior state to the Hot Big Bang expansion phase. The Hartle-Hawking 'no-boundary' proposal [Hawking, 1993] avoids the initial singularity because of a change of space-time signature, and so gets round the issue of a time of creation in an ingenious way; and Gott's causality violation in the early universe [Gott and Li, 1998] does the same kind of thing in a different way. Such proposals cannot overcome the ultimate existential question: *Why has* *one specific state occurred rather than any of the other possibilities? How was it decided that this particular kind of universe would be the one actually instantiated?* This question cannot be solved by physics alone, unless one can show that only one form of physics is self-consistent; but the variety of proposals made is evidence against that suggestion.

The explanation of initial conditions has been the aim of the family of theories one can label collectively as 'quantum cosmology' [Hawking, 1993; Gott and Li, 1998; Gibbons *et al.*, 2003]; however as discussed earlier, here we inevitably reach the limits to what the scientific study of the cosmos can ever say — if we assume that such studies must of necessity involve an ability to observationally or experimentally check our theories. No physical experiment at all can help here because of the uniqueness of the universe, and the feature that no spacetime exists prior to (in a causal sense) such a beginning; so brave attempts to define a 'physics of creation' stretch the meaning of 'physics'. Prior to the start (if there was a start) physics as we know it is not applicable and our ordinary language fails us because time did not exist, so our natural tendency to contemplate what existed or happened 'before the beginning' is highly misleading — there was no 'before' then, indeed there was no 'then' then! Talking as if there was is commonplace, but quite misleading in trying to understand a scientific concept of 'creation' [Grunbaum, 1989]. We run full tilt into the impossibility of testing the causal mechanisms involved, when physics did not exist. No experimental test can determine the nature of any mechanisms that may be in operation in circumstances where even the concepts of cause and effect are suspect. This comes particularly to the fore in proposing 'laws of initial conditions for the universe' — for here we are apparently proposing a theory with only one object. Physics laws are by their nature supposed to cover more than one event, and are untestable if they do not do so (Sec. 3).

6.3 Special or general

The present state of the universe is very special. Explanation of the present large-scale isotropy and homogeneity of the universe means determining the dynamical evolutionary trajectories relating initial to final conditions, and then essentially either *explaining initial conditions*, where we run into difficulties (Sec. 6.2), or *showing they are irrelevant*. The issue raised is whether the universe started off in a very special geometrical state:

Thesis D3: The initial state of the universe may have been special or general. *Whether there was generality or speciality of geometrical initial conditions for the universe is a key question. It seems likely that the initial state of the observed part of the universe was not generic.*

The assumption that the universe is geometrically special was encoded in the Cosmological Principle, taken as a founding principle in cosmology until the 1960's, i.e. as an 'explanation' of special initial conditions [Bondi, 1960; Weinberg, 1972]. Then Misner introduced the chaotic cosmology programme [Misner, 1968], based on the idea of a universe with generic initial conditions being isotropised at later times by physical processes such as viscosity, making initial conditions irrelevant. This concept of isotropisation then became central to the inflationary family of theories (Sec. 2.6), with the underlying assumption being that 'fine tuning' of initial conditions is unphysical and to be avoided. Both programmes are however only partially successful: one can explain a considerable degree of isotropisation and homogenization of the physical universe by either process, but this will not work in all circumstances. Inflation can get rid of much anisotropy [Wald, 1983] but inhomogeneity must be restricted if inflation is to succeed in producing a universe like that we see today, and the success of inflation in solving the horizon issue for FL models — where exact homogeneity exists to start with — will not necessarily be replicated in anisotropic models. Universes that are initially too anisotropic may never inflate, and the horizon problem may not be solved in such models if they do;[26] and only rather special states lead to ordinary thermodynamics [Penrose, 1989a; Penrose, 2004; Wald, 2005; Carroll and Chen, 2005], which is taken to be true in inflationary physics.

Inflation can only be guaranteed to succeed if initial conditions are somewhat restricted; some degree of geometric speciality must have occurred at the start of the observed region of the universe. This special domain might possibly occur within the context of a much larger universe domain where conditions vary randomly, and only isolated regions lead to inflation and eventually domains such as that we see around us; attractive as this may be, it is an untestable hypothesis (essentially a version of the multiverse proposal, see Sec. 9.2).

Special initial conditions (which inflation proposes to explain) might have just occurred that way. The ultimate issue is that *we have no proof as to whether initial conditions for the universe were special or general; either could have occurred.* If we state these conditions must have been general, we are making a philosophical claim, for it is not a provable physical statement. Part of the problem is that we have no agreed measure on the space of possible universes; what seems special or general depends on the choice of such a measure.

7 ISSUE E: THE UNIVERSE AS THE BACKGROUND FOR EXISTENCE

The universe provides the environment for all of science, by determining the initial conditions within which all physical laws are constrained to operate, thus setting boundary conditions for all local physics. Together with suitable equations of state for the matter or structural equations for complex systems, these determine the nature of physical outcomes. The uniqueness of cosmology lies in that it considers the origin of such conditions.

[26] Most inflationary studies show only that the *geometric* horizon problem is solved in the very special RW geometries; but there is no *physical* horizon problem in those geometries, for they are by by assumption spatially homogeneous and isotropic *ab initio*.

7.1 Laws and boundary conditions

A fundamental assumption underlying physical cosmology is the idea that *the laws of physics are the same everywhere in the physical universe*: those we determine in a laboratory here and now will be the same as apply at very distant places (e.g. determining the astrophysics of qso's at redshift $z = 6$), at very early times (e.g. at the time of nucleosynthesis), and at very late times. Without this assumption, explanatory theories have no solid foundation. However because of the uniqueness of the universe discussed above (see Sec. 3), unlike the rest of physics where the distinction is clear and fundamental, in the cosmological context the distinction between laws and boundary conditions becomes blurred.

Thesis E1: Physical laws may depend on the nature of the universe. *We have an essential difficulty in distinguishing between laws of physics and boundary conditions in the cosmological context of the origin of the universe. Effective physical laws may depend on the boundary conditions of the universe, and may even vary in different spatial and/or temporal locations in the cosmos.*

Because we cannot vary the initial conditions in any way, as far as we are concerned they are necessary rather than contingent — so the essential distinction between initial conditions and laws is missing. The distinction is clear once the cosmos has come into existence — but we are concerned with 'prior' conditions associated with the creation of the cosmos and the very existence of physical laws. Certainly any proposal for distinguishing between laws of nature and boundary conditions governing solutions to those laws is untestable in this context. Given the feature that the universe is the unique background for all physics, it is therefore not far-fetched to suggest that it is possible the cosmos influences the *nature* of local physical laws, rather than just their initial conditions [Ellis and Sciama, 1972; Ellis, 2002]. This has been examined over many decades in three specific cases.

(a) *Varying 'constants'*: It might be that there is a time variation in physical constants of nature [Barrow, 2003] related to the expansion of the universe, as proposed in the case of the gravitational constant G by Dirac [Dirac, 1938], developed in depth by Jordan and then Brans and Dicke [Brans and Dicke, 1961]. Such proposals must be consistently developed in relation to the rest of physics and should be related to dimensionless constants, as otherwise they may simply be disguised variations in the units of measurements used, rather than being a genuine physical change (various claims that the speed of light 'c' may vary fall into this category [Ellis and Uzan, 2005]). This proposal has received impetus in recent times from ideas based in quantum field theory and string theory, suggesting that many of the 'constants of nature' are in fact contingent, depending on the nature of the vacuum state [Susskind, 2003; Freivogel *et al.*, 2005a]. This kind of proposal is to some degree open to observational test [Cowie and Songaila, 1995; Will, 1979], and in the cases where it has been investigated it seems that it does not occur in the visible region of the universe — the constants of nature are indeed invariant, with one possible exception: the fine structure constant, where there is claimed to be evidence of a very small change over astronomical timescales [Barrow, 2003]. That issue is still under investigation. Testing such invariance is fundamentally important, precisely because cosmology usually assumes as a ground rule that physics is the same everywhere in the universe. If this were not true, local physics would not guide us adequately as to the behaviour of matter elsewhere or at other times, and cosmology would become an arbitrary guessing game. In order to proceed in a scientific manner when such variation is proposed, one needs then to hypothesize the manner of such variation. Thus the old laws where G was constant are replaced by new laws governing its time variation [Brans and Dicke, 1961]; the principle of nature being governed by invariant (unchanging) physical laws and associated constants remains.[27] Thus in the end the proposal is to replace simpler old laws by new more complex ones. These must then be assumed invariant, or we cannot proceed scientifically.

(b) *Inertia and Mach's Principle*: It might be that the local inertial properties of matter are determined by the distant distribution of matter in the universe, so that if the universe were different, inertia would be different. This is the complex of ideas referred to as Mach's principle [Barbour and Pfister, 1995], which served as a major impetus for Einstein's cosmological ideas. The precise meaning and implications of this idea remain controversial.

(c) *The arrow of time*: The existence and direction of the macroscopic arrow of time in physics — and hence in chemistry, biology, psychology, and society — is related to boundary conditions in the past and future of the universe. The fundamental physical laws by themselves are time symmetric, and so unable to explain this feature [Davies, 1974; Ellis and Sciama, 1972; Zeh, 1992; Uffink, 2006]. A recent argument of this kind is Penrose's claim that the existence of the arrow of time is crucially based in the universe having had rather special initial conditions [Penrose, 1989b; Penrose, 1989a; Wald, 2005]. Thus what appears in ordinary physics as an immutable law of nature (viz. the Second Law of Thermodynamics with a given arrow of time) may well be the result of specific boundary conditions at the start and end of the universe. It might not be true in all universes, even if the underlying fundamental physical laws are the same.

In each case proposals have been made as to the possible nature of the deeper underlying unchanging laws, and the relations between the state of the universe and the resultant effective laws in that context. This is also proposed in the 'landscape' of possibilities of string theory [Susskind, 2005]. These proposals are however intrinsically untestable, for the reasons explained above (Sec. 3): we cannot change

[27] "Despite the incessant change and dynamic of the visible world, there are aspects of the fabric of the universe which are mysterious in their unshakeable constancy. It is these mysterious unchanging things that make our universe what it is and distinguish it from other worlds we might imagine" (Barrow [Barrow, 2003], p. 3).

the boundary conditions of the universe and see what happens; but they do serve as a continuing fertile source of ideas.

7.2 Alternative physics

In any case, the important conclusion is that it is certainly appropriate for cosmology to consider what would have happened if, not only the boundary conditions at the beginning of the universe, but also the laws of physics had been different [Susskind, 2005]:

Thesis E2: We cannot take the nature of the laws of physics for granted. *Cosmology is interested in investigating hypothetical universes where the laws of physics are different from those that obtain in the real universe in which we live — for this may help us understand why the laws of physics are as they are (a fundamental feature of the real physical universe).*

One cannot take the existence and nature of the laws of physics (and hence of chemistry) as unquestionable in cosmology — which seems to be the usual habit in biological discussions on the origin and evolution of life. This is in stark contrast to the rest of science, where we are content to take the existence and nature of the laws describing the fundamental behaviour of matter as given and unchangeable. Cosmological investigation is interested in the properties of hypothetical universes with different physical behaviour. Consideration of 'what might have been' is a useful cosmological speculation that may help throw light on what actually is; this is a statement of the usefulness of 'Gedanken experiments' in cosmology.

Indeed if one wants to investigate issues such as why life exists in the universe, consideration of this larger framework — in essence, a hypothetical ensemble of universes with many varied properties — is essential (this is of course not the same as assuming an ensemble of such universes actually exists, cf. the discussion below in Sec. 9.2). However we need to be very cautious about using any claimed statistics of universes in such a hypothetical ensemble of all possible or all conceivable universes. This is usually not well defined, and in any case is only relevant to physical processes if either the ensemble actually exists, rather than being a hypothetical one, or if it is the outcome of processes that produce well-defined probabilities — an untestable proposal. We can learn from such considerations the nature of possible alternatives, but not necessarily the probability with which they might occur (if that concept has any real meaning).

7.3 Emergence of complexity

As the universe evolves an increase of complexity takes place in local systems as new kinds of objects come into being that did not exist before — nuclei, atoms, stars and galaxies, planets, life, consciousness, and products of the mind such as books and computers [Morowitz, 2002]. New kinds of physical states come into being at late times such as Bose-Einstein condensates, that plausibly cannot exist without the intervention of intelligent beings.

Thesis E3: Physical novelty emerges in the expanding universe. *New kinds of physical existence come into being in the universe as it evolves, that did not exist previously. Their existence is allowed by the boundary conditions provided by the universe for local systems, together with the possibility space generated by the underlying physics. While their physical existence is novel, every new thing that comes into being is foreshadowed in possibility structures that precede their existence.*

Physical existence is new as the universe evolves, but there had to be precursors of the novel in the possibility space allowed by physics, so that they could come into being. In this sense the truly novel does not emerge *ex nihilo* but rather is discovered. The universe is the environment that allows this to happen. The nature of the features leading to the existence of life, and their possible causes, is discussed in Sec. 9.1.

8 ISSUE F: THE EXPLICIT PHILOSOPHICAL BASIS

Consequent on the discussion above, and particularly items **B6**, **C2**, and **D2**, it follows that

Thesis F1: Philosophical choices necessarily underly cosmological theory. *Unavoidable metaphysical issues inevitably arise in both observational and physical cosmology. Philosophical choices are needed in order to shape the theory.*

There is of course always a philosophical basis to any scientific analysis, namely adoption of the basic scientific method and a commitment to the attempt to explain what we see as far as possible simply in terms of causal laws, ultimately based in physics. This will clearly be true also in cosmology. However we need further explicit philosophical input in order to attain specific geometric models — for example a Copernican principle, as explained above, see Sec. 4.2.2 — and to determine what form physical cosmology should take in the very early universe, for example deciding which physical principle to use as the core of one's extrapolation of known physics to the unknown (Sec. 5). Underlying both sets of choices are criteria for satisfactoriness of a cosmological model, which help decide which feature to focus on in formulating a theory. Of particular importance is the scope chosen for our cosmological theory; together with the choice of criteria for a good theory, this is a philosophical decision that will shape the rest of the analysis. Some cosmologists tend to ignore the philosophical choices underlying their theories; but simplistic or unexamined philosophical standpoints are still philosophical standpoints!

8.1 Criteria for theories

As regards criteria for a good scientific theory [Kuhn, 1977], typical would be the following four areas of assessment:

1. *Satisfactory structure*: **(a)** internal consistency, **(b)** simplicity (Occam's razor), and **(c)** aesthetic appeal ('beauty' or 'elegance').

2. *Intrinsic explanatory power*: **(a)** logical tightness, **(b)** scope of the theory — the ability to unify otherwise separate phenomena, and **(c)** probability of the theory or model with respect to some well-defined measure;

3. *Extrinsic explanatory power, or relatedness*: **(a)** connectedness to the rest of science, **(b)** extendability — providing a basis for further development;

4. *Observational and experimental support*, in terms of **(a)** testability: the ability to make quantitative as well as qualitative predictions that can be tested; and **(b)** confirmation: the extent to which the theory is supported by such tests as have been made.

It is particularly the latter that characterizes a scientific theory, in contrast to other types of theories claiming to explain features of the universe and why things happen as they do. It should be noted that *these criteria are philosophical in nature in that they themselves cannot be proven to be correct by any experiment.* Rather their choice is based on past experience combined with philosophical reflection. One could attempt to formulate criteria for good criteria for scientific theories, but of course these too would need to be philosophically justified. The enterprise will end in infinite regress unless it is ended at some stage by a simple acceptance of a specific set of criteria.

Thesis F2: Criteria of satisfactoriness for theories cannot be scientifically chosen or validated. *Criteria of satisfactoriness are necessary for choosing good cosmological theories; these criteria have to be chosen on the basis of philosophical considerations. They should include criteria for satisfactory structure of the theory, intrinsic explanatory power, extrinsic explanatory power, and observational and experimental support.*

The suggestion here is that the above proposed criteria are a good set to use in investigating cosmology; they include those most typically used ([Kuhn, 1977]; and see [Penrose, 2004; Susskind, 2005] for comments on such criteria).

8.1.1 Conflicts between criteria.

These criteria are all acknowledged as desirable. The point then is that generally in pursuing historical sciences, and in particular in the cosmological context, they will not all be satisfied to the same degree, and may even lead to opposing conclusions:

Thesis F3: Conflicts will inevitably arise in applying criteria for satisfactory cosmological theories. *Philosophical criteria for satisfactory cosmological theories will in general come into conflict with each other, so that one will have to choose between them to some degree; this choice will shape the resulting theory.* [Ellis, 1991].

The thrust of much recent development has been away from observational tests towards strongly theoretically based proposals, indeed sometimes almost discounting observational tests. At present this is being corrected by a healthy move to detailed observational analysis of the consequences of the proposed theories, marking a maturity of the subject. However because of all the limitations in terms of observations and testing [criteria **(4)**], in the cosmological context we still have to rely heavily on other criteria, and some criteria that are important in most of science may not really make sense. This is true of **2(c)** in particular, as discussed above, see Sec. 3; nevertheless many approaches still give the idea of probability great weight. At a minimum, the ways this can make sense needs exploration and explication. Furthermore the meaning of some of the criteria may come into dispute. **1(b)** is clearly a case in point : for example, is the idea of an existent ensemble of universes displaying all possible behaviours simple (because it is a single idea that can be briefly stated), or immensely complex (because that statement hides all the complexities and ambiguities involved in the idea of an infinity of possibilities)? **1(c)** is also controversial('beauty is in the eye of the beholder'), see [Susskind, 2005] for a discussion.

The tenor of scientific understanding may change, altering the balance of what is considered a good explanation and what is not. An example [Ellis, 1990] is the way cosmologists strongly resisted the idea of an evolving universe in the 1920's, at a time when biological evolution was very well established but the idea of continental drift was also being strongly resisted. The change to an appreciation of the explanatory power of an evolving model came later in both cases; but even then in the cosmological case, for either aesthetic or metaphysical reasons, some still sought for a steady state description, resisting the implication of a beginning to the universe. That tendency is still with us today, in the form of models that are eternal in one way or another (e.g. some forms of chaotic inflation). Another example is the change from supposition of underlying order, expressed in the idea of a Cosmological Principle, to a broad supposition of generic disordered conditions, embodied in the ideas of inflation. Associated with this is a shift from making geometric assumptions to providing physical explanatory models. It is this shift that underlies the major present support for inflation:

Thesis F4: The physical reason for believing in inflation is its explanatory power as regards structure growth in the universe. *Inflation predicts the existence of Gaussian scale-free perturbations in the early universe thereby (given the presence of cold dark matter) explaining bottom-up structure formation in a satisfactory way. This theory has been vindicated spectacularly through observations of the CBR and matter power spectra. It is this explanatory power that makes it so acceptable to physicists, even though the underlying physics is neither well-defined nor tested, and its major large-scale observational predictions are untestable.*

The physical explanatory power of inflation in terms of structure formation, supported by the observational data on the fluctuation spectra, is spectacular. For most physicists, this trumps the lack of identification and experimental verification of the underlying physics (Sec. 5). Inflation provides a causal model that brings a wider range of phenomena into what can be explained by cosmology (Criterion **2(b)**), rather than just assuming the initial data had a specific restricted form. Explaining flatness ($\Omega_0 \simeq 1$ as predicted by inflation) and homogeneity reinforces the case, even though these are philosophical rather than physical problems (they do not contradict any physical law; things could just have been that way). However claims on the basis of this model as to what happens very far outside the visual horizon (as in the chaotic inflationary theory) results from prioritizing theory over the possibility of observational and experimental testing [Earman and Mosterin, 1999]. It will never be possible to *prove* these claims are correct.

8.2 The scope of cosmology

To sensibly choose priorities for the criteria just discussed, we need an answer to the question, How much should we try to explain?

Thesis F5: Cosmological theory can have a wide or narrow scope of enquiry. *The scope we envisage for our cosmological theory shapes the questions we seek to answer. The cosmological philosophical base becomes more or less dominant in shaping our theory according to the degree that we pursue a theory with more or less ambitious explanatory aims in terms of all of physics, geometry, and underlying fundamental causation.*

This is a choice one has to make, as regards both foundations and outcomes. Given a decision on this, one can sensibly debate what is the appropriate philosophical position to adopt in studying a cosmological theory with that scope. The study of expansion of the universe and structure formation from nucleosynthesis to the present day is essential and well-informed. The philosophical stance adapted is minimal and highly plausible. The understanding of physical processes at earlier times, back to quantum gravity, is less well founded. The philosophical stance is more significant and more debatable. Developments in the quantum gravity era are highly speculative; the philosophical position adopted is dominant because experimental and observational limits on the theory are lacking.

One can choose the degree to which one will pursue the study of origins [Fabian, 1989] back to earlier and earlier times and to more fundamental causal issues, and hence the degree to which specific philosophical choices are dominant in one's theory. The basic underlying cosmological questions are [Ellis, 1991]:

1. *Why do the laws of physics have the form they do?* Issues arise such as what makes particular laws work? For example, what guarantees the behaviour of a proton, the pull of gravity? What makes one set of physical laws 'fly' rather than another? If for example one bases a theory of cosmology on string theory [Susskind, 2005], then who or what decided that quantum gravity would have a nature well described by string theory? If one considers

all possibilities, considering string theory alone amounts to a considerable restriction.

2. *Why do boundary conditions have the form they do ?* The key point here (Sec. 6.2), is how are specific contingent choices made between the various possibilities, for example whether there was an origin to the universe or not.

3. *Why do any laws of physics at all exist ?* This relates to unsolved issues concerning the nature of the laws of physics: are they descriptive or prescriptive? (Sec. 9.3.3). Is the nature of matter really mathematically based in some sense, or does it just happen that its behaviour can be described in a mathematical way?

4. *Why does anything exist ?* This profound existential question is a mystery whatever approach we take.[28]

   Finally the adventurous also include in these questions the more profound forms of the contentious Anthropic question [Carr and Rees, 1979; Davies, 1982; Barrow and Tipler, 1984; Tegmark, 1998; Susskind, 2005]:

5. *Why does the universe allow the existence of intelligent life?* This is of somewhat different character than the others and largely rests on them but is important enough to generate considerable debate in its own right.

The status of all these questions is philosophical rather than scientific, for they cannot be resolved purely scientifically. How many of them — if any — should we consider in our construction of and assessments of cosmological theories?

One option is *to decide to treat cosmology in a strictly scientific way*, excluding all the above questions, because they cannot be solved scientifically. One ends up with a solid technical subject that by definition excludes such philosophical issues. This is a consistent and logically viable option. This logically unassailable position however has little explanatory power; thus most tend to reject it because of criteria **2(b)** and **3** above.

The second option is to decide that *these questions are of such interest and importance that one will tackle some or all of them, even if that leads one outside the strictly scientific arena.* It is here that criteria **2** and **3** above are to some degree in conflict with criterion **4**. Thus if we try to explain the origin of the universe itself, these philosophical choices become dominant precisely because the experimental and observational limits on the theory are weak; this can be seen by viewing the variety of such proposals that are at present on the market.

8.3 Limits of Representation and Knowledge of Reality

It follows from the above discussion that there are limits to what the scientific method can achieve in explanatory terms. We need to respect these limits and

[28] But see Grunbaum [Grunbaum, 2004] for a dissenting view.

acknowledge clearly when arguments and conclusions are based on some philosophical stance rather than purely on testable scientific argument. If we acknowledge this and make that stance explicit, then the bases for different viewpoints are clear and alternatives can be argued about rationally.

A crucial underlying feature here is relating the nature of epistemology to ontology: how do we relate evidence to our theories of existence? A further key issue is the relation of models to reality:

Thesis F6: Reality is not fully reflected in either observations or theoretical models. *Problems arise from confusion of epistemology (the theory of knowledge) with ontology (the nature of existence): existence is not always manifest clearly in the available evidence. The theories and models of reality we use as our basis for understanding are necessarily partial and incomplete reflections of the true nature of reality, helpful in many ways but also inevitably misleading in others. They should not be confused with reality itself!*

The confusion of epistemology with ontology occurs all the time, underlying for example the errors of both logical positivism and extreme relativism. In particular, it is erroneous to assume that lack of evidence for the existence of some entity is proof of its non-existence. In cosmology it is clear for example that regions may exist from which we can obtain no evidence (because of the existence of horizons); so we can sometimes reasonably deduce the existence of unseen matter or regions from a sound extrapolation of available evidence (no one believes matter ends at or just beyond the visual horizon). However one must be cautious about the other extreme, assuming existence can always be assumed because some theory says so, regardless of whether there is any evidence of existence or not. This happens in present day cosmology, for example in presentations of the case for multiverses, even though the underlying physics has not been experimentally confirmed. It may be suggested that arguments ignoring the need for experimental/observational verification of theories ultimately arise because these theories are being confused with reality, or at least are being taken as completely reliable total representations of reality. This occurs in

- Confusing computer simulations of reality with reality itself, when they can in fact represent only a highly simplified and stylized version of what actually is;
- Confusing the laws of physics themselves with their abstract mathematical representation (if indeed they are ontologically real, c.f. Sec. 10.1), or confusing a construction of the human mind ('Laws of Physics') with the reliable behaviour of ponderable matter (if they are not ontologically real);
- Confusing theoretically based outcomes of models with proven observational results (e.g. claiming the universe necessarily has flat spatial sections: $\Omega_0 = 1$, and so this can be taken for granted, when the value of Ω_0 can and should be observationally determined precisely because this then tests that prediction).

No model (literary, intuitive, or scientific) can give a perfect reflection of reality. Such models are always selective in what they represent and partial in the completeness with which they do so. The only model that would reflect reality fully is a perfect fully detailed replica of reality itself! This understanding of the limits of models and theories does not diminish the utility of these models; rather it helps us use them in the proper way. This is particularly relevant when we consider how laws of nature may relate to the origins of the universe itself, and to the existence and nature of life in the expanding universe. The tendency to rely completely on our theories, even when untested, seems sometimes to arise because we believe they are the same as reality — when at most they are *descriptions* of reality.

9 KEY ISSUES

There are some interrelated key issues where the features identified above either are at the heart of current debates, or are likely to be at the heart of future debates. They are: the reason cosmological conditions allow the existence of life (anthropic issues), the closely related issue of the possible existence of multiverses; and the natures of existence, including the questions of the existence of infinities and the nature of the laws of physics. We look at them in turn in this section. To some degree they have already been considered above, but they are specifically featured here because of the important role they will probably play in discussion in the future.

9.1 Issue G: The anthropic question: Fine tuning for life

One of the most profound fundamental issues in cosmology is the Anthropic question, see [Davies, 1982; Barrow and Tipler, 1984; Earman, 1987; Fabian, 1989; Davies, 1987; Balashov, 1991; Rees, 1999; Rees, 2003; Barrow, 2003]: *why does the Universe have the very special nature required in order that life can exist?*. The point is that a great deal of "fine tuning" is required in order that life be possible. There are many relationships embedded in physical laws that are not explained by physics, but are required for life to be possible; in particular various fundamental constants are highly constrained in their values if life as we know it is to exist:

"A universe hospitable to life — what we might call a biophilic universe — has to be special in many ways ... Many recipes would lead to stillborn universes with no atoms, no chemistry, and no planets; or to universes too short lived or too empty to evolve beyond sterile uniformity" [Rees, 2003].

How has it come about that the Universe permits the evolution and existence of intelligent beings at any time or place? "What features of the universe were essential for creatures such as ourselves, and is it through coincidence or for some deeper reason that our universe has these features?" [Gribbin and Rees, 1991]. Whether one regards this as an appropriate issue for cosmology to discuss depends, as discussed above (Sec. 8.2), on the scope one envisages for cosmology. The viewpoint taken here will be that this is one of the major issues one might wish to explain, and indeed a substantial literature considers this. Here we explore the nature of this fine tuning, and then consider possible answers as to how it arises. There are three aspects that we consider in turn (cf. [Susskind, 2005]).

9.1.1 Laws of physics and the existence of complexity

The laws of physics and chemistry are such as to allow the functioning of living cells, individuals, and ecosystems of incredible complexity and variety, and it is this that has made evolution possible. What requires explanation, *is why the laws of physics are such as to allow this complex functionality to work*, without which no evolution whatever would occur. We can conceive of universes where the laws of physics (and so of chemistry) were different than in ours. Almost any change in these laws will prevent life as know it from functioning.

The first requirement is *the existence of laws of physics that guarantee the kind of regularities that can underlie the existence of life*. These laws as we know them are based on variational and symmetry principles; we do not know if other kinds of laws could produce complexity. If the laws are in broad terms what we presently take them to be, the following *inter alia* need to be right, for life of the general kind we know to exist [Davies, 1982; Gribbin and Rees, 1991]:

- Quantization that stabilizes matter and allows chemistry to exist through the Pauli exclusion principle.
- The neutron-proton mass differential must be highly constrained. If the neutron mass were just a little less than it is, proton decay could have taken place so that by now no atoms would be left at all [Davies, 1982].
- Electron-proton charge equality is required to prevent massive electrostatic forces overwhelming the weaker electromagnetic forces that govern chemistry.
- The strong nuclear force must be strong enough that stable nuclei exist [Davies, 1982]; indeed complex matter exists only if the properties of the nuclear strong force lies in a tightly constrained domain relative to the electromagnetic force [Tegmark, 2003].
- The chemistry on which the human body depends involves intricate folding and bonding patterns that would be destroyed if the fine structure constant (which controls the nature of chemical bonding) were a little bit different.
- The number D of large spatial dimensions must be just 3 for complexity to exist [Tegmark, 2003; Rees, 2003].

Hogan has examined the freedom in the parameters of the standard model of particle physics and concluded that 5 of the 17 free parameters of the standard model must lie in a highly constrained domain if complex structures are to exist [Hogan, 2003]. This is of course taking the basic nature of the standard model of particle physics for granted. If this were not so, it is difficult to determine what the constraints would be. However his study is sufficient to show that whatever the nature of fundamental physics, and in particular of particle physics, may be, only a small subset of all possible laws of physics will be compatible with the existence of complexity.

9.1.2 Laws of physics and the existence of congenial environments

The creation through astrophysical processes of suitable habitats for life to exist (the existence of planets circling stable stars, for example) depends to some degree on the nature of the fundamental physical laws. If the laws are in broad terms what we presently take them to be, the requirements for such habitats to exist include:

- The gravitational force must create large stable structures (planets and stars) that can be the habitat for life and their energy source respectively. This requires the gravitational force to be very weak relative to electrical forces. The ratio $\mathcal{N}$ of the strength of the electromagnetic force to the gravitational force must be close to the observed value: $\mathcal{N} \simeq 10^{36}$ [Rees, 1999, Ch. 3].
- The weak force must allow helium production that leaves sufficient hydrogen over; it is related to gravity through a numerical factor of 10^{-11}, which cannot be much different. And for this to work, the neutron-proton mass difference must be close to the mass of the electron [Davies, 1982].
- A stellar balance should allow a long lifetime for stars like the sun, so allowing the transmutation of the light elements into heavy elements. This requires that the nuclear fusion efficiency $\mathcal{E}$ be close to the observed value: $\mathcal{E} \simeq 0.007$ [Rees, 1999, Ch. 4].
- One needs to overcome the beryllium "bottleneck" in the making of heavy elements through nuclear reactions in stars [Gribbin and Rees, 1991; Susskind, 2005]. The production of carbon and oxygen in stars requires the careful setting of two different nuclear energy levels to provide a resonance; if these levels were just a little different, the elements we need for life would not exist [Fabian, 1989]. Indeed it was on this basis that Hoyle famously predicted a carbon-12 energy level that has since been experimentally confirmed.
- One needs something like the existence of neutrinos and the weak interaction with its specific coupling constant in order to underly supernovae explosions that spread heavy elements through space, as seeds for planetary formation [Gribbin and Rees, 1991].
- The nuclear force must be weak enough that di-protons do not exist, otherwise no protons will be left over to enable heavier elements to exist [Davies, 1982].
- The neutrino mass must not be too high, or the universe will not last long enough [Davies, 1982].

9.1.3 Cosmological boundary/initial conditions and congenial environments

Finally, given laws of physics that are suitable in terms of satisfying the requirements of both the previous sections, the universe itself must also be suitable, in terms of its initial or boundary conditions, for life to exist. If the laws of physics are basically the same as we now believe them to be, these cosmological requirements include

- The size of the universe and its age must be large enough. There could be universes that expanded and then recollapsed with a total lifetime of only 100,000 years; we need a sufficiently old universe for second generation stars to come into existence and then for planets to have a stable life for long enough that evolution could lead to the emergence of intelligent life. Thus the universe must be at about 15 billion years old for life to exist [Gribbin and Rees, 1991], hence we must have $\Omega_{matter} \simeq 0.3$ [Rees, 1999, Ch. 6].

- The size of the cosmological constant must not be too large, or galaxies will not form; we need $|\Omega_\Lambda| < 1$ for galaxies to exist [Rees, 1999, Ch. 7]; [Susskind, 2005].

- The seeds in the early universe for fluctuations that will later grow into galaxies must be of the right size that structures form without collapsing into black holes: the number Q characterizing the size of primordial ripples on the LSS (and hence the geometry of the perturbed cosmological model, see Sec. 2.5.2) must therefore be of the order $Q \simeq 10^{-5}$ [Rees, 1999, Ch. 8].

The complex of interacting systems in a human body could not possibly work if a series of delicate conditions were not maintained. For example, the background radiation might never drop below 3000 K, so that matter was always ionized (electrons and nuclei always remaining separate from each other); the molecules of life could then never form. Black holes might be so common that they rapidly attracted all the matter in the universe, and there never was a stable environment in which life could develop. Cosmic rays could always be so abundant that any tentative organic structures were destroyed before they could replicate. Overall,

- There must be non-interference with local systems. The concept of locality is fundamental, allowing local systems to function effectively independently of the detailed structure of the rest of the Universe. We need the universe and the galaxies in it to be largely empty, and gravitational waves and tidal forces to be weak enough,[29] so that local systems can function in a largely isolated way [Ellis, 2002].

- The fact that the night sky is dark ('Olbers' paradox' [Bondi, 1960; Harrison, 2000]) is a consequence of the expansion of the universe together with the photon to baryon ratio. This feature is a necessary condition for the existence

[29]Thus the Weyl tensor C_{abcd} must be suitably small everywhere, presumably implying an almost-RW geometry, cf. [Stoeger *et al.*, 1995].

of life: the biosphere on Earth functions by disposing of waste energy to the heat sink of the dark night sky [Penrose, 1989b]. Thus one way of explaining why the sky is observed to be dark at night is that if this were not so, we would not be here to observe it.

- The existence of the arrow of time, and hence of laws like the second law of thermodynamics, are probably necessary for evolution and for consciousness. This depends on boundary conditions at the beginning and end of the Universe (Sec. 7.1).
- Presumably the emergence of a classical era out of a quantum state is required. The very early universe would be a domain where quantum physics would dominate, leading to complete uncertainty and an inability to predict the consequence of any initial situation; we need this to evolve to a state where classical physics leads to the properties of regularity and predictability that allow order to emerge.
- Physical conditions on planets must be in a quasi-equilibrium state for long enough to allow the delicate balances that enable our existence, through the very slow process of evolution, to be fulfilled.

Thus the existence of suitable local systems to be a habitat for life depends critically on the large-scale properties of very distant matter. These provides a stable local environment within which life can develop.

9.1.4 Fine tuning overall

Thus there are many ways that conditions in a universe could prevent life occurring. Life will occur only if: there exist heavy elements; there is sufficient time for evolution of advanced life forms to take place; there are regions in the universe that are neither too hot nor too cold; there are precisely restricted values of the fundamental constants that control chemistry and local physics; and so on. These conditions will not be true in a generic universe. In summary,

Thesis G1: Life is possible because both the laws of physics and the boundary conditions for the universe have a very special nature. *Only particular laws of physics, and particular initial conditions in the Universe, allow the existence of intelligent life of the kind we know. No evolutionary process whatever is possible for any kind of life if these laws and conditions do not have this restricted form.*

Why is this so? One should note that we can only meaningfully refer here to 'life as we know it'. One of the recurring issues is whether there could be some other quite different basis for life. You can if you wish speculate that life might exist in some immaterial form, or based only on light elements, or existent deep in empty space without the need for stars or planets to provide a viable habitat. The anthropic literature is based on assuming this is not viable, but we cannot *prove* anything in this regard. We have no idea of any basis by which life might come into existence other than the broad principles we see in the life around us. The basic principles of life as we understand it require a great degree of complex organization enabling it to fulfil a complex variety of functions that can only, as far as we know, be based in material existence with information storage, energy usage, sensing of the external world, etc., which requires at a minimum heavy elements (carbon, nitrogen, oxygen, phosphorus for example), a long-term energy source (such as the flow of energy from the sun), and a stable environment (such as the surface of a planet). When we abandon this basis for understanding — saying 'yes but some other form of life might exist' without providing any proposal for its possible structure — one enters the unprofitable realm of speculation. It does not seem to provide any useful way forward.

9.1.4.1 The Weak Anthropic Principle. There are two purely scientific approaches to the Anthropic issue.[30] The first is the *Weak Anthropic Principle* (WAP), based on the comment: it is not surprising the observed Universe admits the existence of life, for the Universe cannot be observed unless there are observers in it [Barrow and Tipler, 1984; Balashov, 1991]. This seemingly empty statement gains content when we turn it round and ask, at what times and places in the Universe can life exist, and what are the inter-connections that are critical for its existence? It could not for example exist too early in the present expansion phase, for the night sky would then have been too hot. Furthermore one can deduce various necessary relations between fundamental quantities in order that the observers should exist (e.g. those mentioned above), so that if for example the fundamental constants vary with time or place in the Universe, life will only be possible in restricted regions where they take appropriate Anthropic values.

Hence this view basically interprets the Anthropic principle as a selection principle: the necessary conditions for observers to exist restricts the times and places from which the Universe can be observed. Because it is quite possible that conditions would not be right for life to exist anywhere in an arbitrarily selected universe, it is also usually conjoined with the idea of the existence of a multiverse, as discussed below, see Sec. 9.2. This is an interesting and often illuminating viewpoint. For example, neither the Chaotic Inflationary Universe idea (Sec. 2.6) nor any other multiverse proposal works unless we add such an Anthropic component into their interpretation to explain why we observe the Universe from a viewpoint where life exists. It is now used by some physicists to explain the low value of the cosmological constant (which quantum field theory predicts should have a very much larger value than observed, see Sec. 9.2.5), and occurs in the context of the possibility landscape of string theory [Susskind, 2005].

[30]I omit the so-called *Final Anthropic Principle* (FAP for short), which maintains that intelligent life must necessarily evolve and then remain in existence until the end of the universe, for I do not believe it merits serious discussion as a scientific proposal; indeed it led to a famous book review referring to the *Completely Ridiculous Anthropic Principle* (CRAP for short) [Gardner, 1986].

9.1.4.2 The Strong Anthropic Principle. By contrast, the *Strong Anthropic Principle* (SAP) [Barrow and Tipler, 1984; Balashov, 1991] claims that it is necessary that intelligent life exist in the Universe; the presence of life is required in order that a universe model make sense. This is clearly a very controversial claim, for it is hard to provide scientific reasons to support this view. One can suggest that the most solid justification attempted is through the claim that the existence of an observer is necessary in order that quantum theory can make sense. However, this justification is based on one of a number of different interpretations of quantum theory; the nature of these quantum foundations is controversial, and not resolved [Isham, 1997; Dickson, 2006; Landsman, 2006].

Furthermore if we were to suppose this justification correct, then the next step is to ask: Why does the Universe need quantum mechanics anyway? The argument would be complete only if we could prove that quantum mechanics was absolutely necessary for every self-consistent Universe; but that line of reasoning cannot be completed at present, not least because quantum mechanics itself is not a fully self-consistent theory. Apart from the conceptual problems at its foundation due to the unresolved measurement issue [Isham, 1997], it suffers from divergences that so far have proved irremediable in the sense that we can work our way round them to calculate what we need, but cannot remove them. The SAP proposal has no accepted physical foundation, and also raises problematic philosophical issues [Earman, 1987]. I will not pursue it further here.

9.1.5 The relation to fundamental physical theories

Many physicists go further, rejecting any Anthropic form of reasoning. They regard it as a cop-out resorted to when physical theories fail to give the needed answers, and seek to obtain a full answer from physics alone [Scott, 2005; Susskind, 2005]. One possibility is that there is a fundamental theory of everything that determines the nature of physics completely, with no arbitrary parameters left, and this still to be discovered theory just happens to be of such a nature as to admit life.

However in this case the Anthropic issue returns with a vengeance: *How could it be that such a theory, based for example on variational principles and the specific invariance groups of particle physics, could just happen to lead to biophilic parameter values*? There is no clear way to answer such a question. Uniqueness of fundamental physics resolves the parameter freedom only at the expense of creating an even deeper mystery, with no way of resolution apparent. In effect, the nature of the unified fundamental force would be pre-ordained to allow, or even encourage, the existence of life; but there would be no apparent reason why this should be so.

A second possibility is that physics allows many effective theories with varying parameters — some form of multiverse, as for example may be implied by string theory [Susskind, 2003; Freivogel *et al.*, 2005a; Susskind, 2005]. If these varying options are all equally real, life can occur because in some cases the parameters will lie in the restricted biophilic regime. Thus from this viewpoint the Anthropic idea is intimately linked with the existence of multiverses, which provide a legitimate domain for their application. We will turn to an examination of multiverses in the next section, but before doing so we will consider the range of metaphysical options for resolving the anthropic question.

9.1.6 The metaphysical options

To make progress on the Anthropic issue, we have to seriously consider the nature of ultimate causation: What is the fundamental cause for the phenomena we see? If we pursue the chain of physical cause and effect to its conclusion, we are still left with the question: *Why did this occur, and not something else*? Whatever the reason is, it is the ultimate cause we are seeking. Note that we are here leaving the terrain of science itself, and starting to probe the domain of metaphysics — the foundations of science and indeed of existence. As noted above, one can simply decide not to pursue such issues. If we do continue to question, there appear to be basically six approaches to the issue of ultimate causation: namely Random Chance, Necessity, High Probability, Universality, Cosmological Natural Selection, and Design. We briefly consider these in turn.

Option 1: *Random Chance, signifying nothing.* The initial conditions in the Universe just happened, and led to things being the way they are now, by pure chance. Probability does not apply. There is no further level of explanation that applies; searching for 'ultimate causes' has no meaning.

This is certainly logically possible, but not satisfying as an explanation, as we obtain no unification of ideas or predictive power from this approach. Nevertheless some implicitly or explicitly hold this view.

Option 2: *Necessity.* Things have to be the way they are; there is no other option. The features we see and the laws underlying them are demanded by the unity of the Universe: coherence and consistency require that things must be the way they are; the apparent alternatives are illusory. Only one kind of physics is self-consistent: all logically possible universes must obey the same physics.

To really prove this would be a very powerful argument, potentially leading to a self-consistent and complete scientific view. But we can imagine alternative universes! — why are they excluded? Furthermore we run here into the problem that we have not succeeded in devising a fully self-consistent view of physics: neither the foundations of quantum physics nor of mathematics are on a really solid consistent basis. Until these issues are resolved, this line cannot be pursued to a successful conclusion.

Option 3: *High probability.* Although the structure of the Universe appears very improbable, for physical reasons it is in fact highly probable.

These arguments are only partially successful, even in their own terms. They run into problems if we consider the full set of possibilities: discussions proposing this kind of view actually implicitly or explicitly restrict the considered possibilities *a priori*, for otherwise it is not very likely the Universe will be as we see it. Besides, we do not have a proper measure to apply to the set of initial conditions, enabling us to assess these probabilities. Furthermore, as discussed above, see Sec. 3, application of probability arguments to the Universe itself is dubious, because the Universe is unique. Despite these problems, this approach has considerable support in the scientific community, for example it underlies the chaotic inflationary proposal (Sec. 2.6). It attains its greatest power in the context of the assumption of universality:

Option 4: *Universality*. This is the stand that "All that is possible, happens": an ensemble of universes or of disjoint expanding universe domains is realized in reality, in which all possibilities occur [Rees, 1999; Rees, 2003; Tegmark, 2003]. In its full version, the anthropic principle is realized in both its strong form (if all that is possible happens, then life must happen) and its weak form (life will only occur in some of the possibilities that are realized; these are picked out from the others by the WAP, viewed as a selection principle). There are four ways this has been pursued.

1. *Spatial variation*. The variety of expanding universe domains is realised in space through random initial conditions, as in chaotic inflation (Sec. 2.6). While this provides a legitimate framework for application of probability, from the viewpoint of ultimate explanation it does not really succeed, for there is still then one unique Universe whose (random) initial conditions need explanation. Initial conditions might be globally statistically homogeneous, but also there could be global gradients in some physical quantities so that the Universe is not statistically homogeneous; and these conditions might be restricted to some domain that does not allow life. It is a partial implementation of the ensemble idea; insofar as it works, it is really a variant of the "high probability" idea mentioned above. If it was the more or less unique outcome of proven physics, then that would provide a good justification; but the physics underlying such proposals is not even uniquely defined, much less tested. Simply claiming a particular scalar field with some specific stated potential exists does not prove that it exists!

2. *Time variation*. The variety of expanding universe domains could be realised across time, in a universe that has many expansion phases (a Phoenix universe), whether this occurs globally or locally. Much the same comments apply as in the previous case.

3. *Quantum Mechanical*. It could occur through the existence of the Everett-Wheeler "many worlds" of quantum cosmology, where all possibilities occur through quantum branching [Deutsch, 1998]. This is one of the few genuine alternatives proposed to the Copenhagen interpretation of quantum mechanics, which leads to the necessity of an observer, and so potentially to the

Strong Anthropic interpretation considered above (see Sec. 9.1). The many-worlds proposal is controversial: it occurs in a variety of competing formulations [Isham, 1997], none of which has attained universal acceptance. The proposal does not provide a causal explanation for the particular events that actually occur: if we hold to it, we then have to still explain the properties of the particular history we observe (for example, why does our macroscopic universe have high symmetries when almost all the branchings will not?). And above all it is apparently untestable: there is no way to experimentally prove the existence of all those other branching universes, precisely because the theory gives the same observable predictions as the standard theory.

4. *Completely disconnected.* They could occur as completely disconnected universes: there really is an ensemble of universes in which all possibilities occur, without any connection with each other [Lewis, 1986; Rees, 2003; Tegmark, 2003]. A problem that arises then is, What determines what is possible? For example, what about the laws of logic themselves? Are they inviolable in considering all possibilities? We cannot answer, for we have no access to this multitude of postulated worlds. We explore this further below (Sec. 9.2).

In all these cases, major problems arise in relating this view to testability and so we have to query the meaningfulness of the proposals as scientific explanations. They all contradict Occam's razor: we "solve" one issue at the expense of envisaging an enormously more complex existential reality. Furthermore, they do not solve the ultimate question: *Why does this ensemble of universes exist*? One might suggest that ultimate explanation of such a reality is even more problematic than in the case of single universe. Nevertheless this approach has an internal logic of its own which some find compelling. We consider this approach further below, see Sec. 9.2.

Option 5: *Cosmological Natural Selection.* If a process of re-expansion after collapse to a black hole were properly established, it opens the way to the concept not merely of evolution of the Universe in the sense that its structure and contents develop in time, but in the sense that the Darwinian selection of expanding universe regions could take place, as proposed by Smolin [Smolin, 1992]. The idea is that there could be collapse to black holes followed by re-expansion, but with an alteration of the constants of physics through each transition, so that each time there is an expansion phase, the action of physics is a bit different. The crucial point then is that some values of the constants will lead to production of more black holes, while some will result in less. This allows for evolutionary selection favouring the expanding universe regions that produce more black holes (because of the favourable values of physical constants operative in those regions), for they will have more "daughter" expanding universe regions. Thus one can envisage natural selection favouring those physical constants that produce the maximum number of black holes.

The problem here is twofold. First, the supposed 'bounce' mechanism has never been fully explicated. Second, it is not clear — assuming this proposed process can be explicated in detail — that the physics which maximizes black hole production is necessarily also the physics that favours the existence of life. If this argument could be made water-tight, this would become probably the most powerful of the multiverse proposals.

Option 6: *Purpose or Design.* The symmetries and delicate balances we observe require an extraordinary coherence of conditions and cooperation of causes and effects, suggesting that in some sense they have been purposefully designed. That is, they give evidence of intention, both in the setting of the laws of physics and in the choice of boundary conditions for the Universe. This is the sort of view that underlies Judaeo-Christian theology. Unlike all the others, it introduces an element of meaning, of signifying something. In all the other options, life exists by accident; as a chance by-product of processes blindly at work.

The prime disadvantage of this view, from the scientific viewpoint, is its lack of testable scientific consequences ("Because God exists, I predict that the density of matter in the Universe should be x and the fine structure constant should be y"). This is one of the reasons scientists generally try to avoid this approach. There will be some who will reject this possibility out of hand, as meaningless or as unworthy of consideration. However it is certainly logically possible. The modern version, consistent with all the scientific discussion preceding, would see some kind of purpose underlying the existence and specific nature of the laws of physics and the boundary conditions for the Universe, in such a way that life (and eventually humanity) would then come into existence through the operation of those laws, then leading to the development of specific classes of animals through the process of evolution as evidenced in the historical record. Given an acceptance of evolutionary development, it is precisely in the choice and implementation of particular physical laws and initial conditions, allowing such development, that the profound creative activity takes place; and this is where one might conceive of design taking place.[31]

However from the viewpoint of the physical sciences *per se*, there is no reason to accept this argument. Indeed from this viewpoint there is really no difference between design and chance, for they have not been shown to lead to different physical predictions.

9.1.7 Metaphysical Uncertainty

In considering ultimate causation underlying the anthropic question, in the end we are faced with a choice between one of the options above. As pointed out already by Kant and Hume, although we may be able to argue strongly for one or other of them, we cannot *prove* any of the options are correct [Hume, 1993].

[31] This is not the same as the view proposed by the 'Intelligent Design' movement. It does not propose that God tweaks the outcome of evolutionary processes.

Thesis G2: Metaphysical uncertainty remains about ultimate causation in cosmology. *We cannot attain certainty on the underlying metaphysical cosmological issues through either science or philosophy.*

If we look at the anthropic question from a purely scientific basis, we end up without any resolution, basically because science attains reasonable certainty by limiting its considerations to restricted aspects of reality; even if it occasionally strays into the area of ultimate causation, it is not designed to deal with it. By itself, it cannot make a choice between these options; there is no relevant experiment or set of observations that can conclusively solve the issue. Thus a broader viewpoint is required to make progress, taking into account both the scientific and broader considerations. The issue is of a philosophical rather than scientific nature. One important issue that then arises is what kind of data is relevant to these philosophical choices, in addition to that which can be characterized as purely scientific data (Sec. 9.3.4).

9.2 Issue H: The possible existence of multiverses

If there is a large enough ensemble of numerous universes with varying properties, it may be claimed that it becomes virtually certain that some of them will just happen to get things right, so that life can exist; and this can help explain the fine-tuned nature of many parameters whose values are otherwise unconstrained by physics [Rees, 1999; Rees, 2003]. As discussed in the previous section, there are a number of ways in which, theoretically, multiverses could be realized [Lewis, 1986; Tegmark, 2003]. They provide a way of applying probability to the universe [Sciama, 1971; Bostrom, 2002] (because they deny the uniqueness of the universe). However, there are number of problems with this concept. Besides, this proposal is observationally and experimentally untestable; thus its scientific status is debatable.

9.2.1 Definition

In justifying multiverses, it is often stated that 'all that can occur, occurs' (or similarly). However that statement does not adequately specify a multiverse. To define a multiverse properly requires two steps [Ellis *et al.*, 2004]. First, one needs to specify what is conceived of in the multiverse, by defining a *possibility space*: a space $\mathcal{M}$ of all possible universes, each of which can be described in terms of a set of states s in a state space $\mathcal{S}$. Each universe m in $\mathcal{M}$ will be characterized by a set of distinguishing parameters p, which are coordinates on $\mathcal{S}$. Choices are needed here. In geometrical terms, will it include only Robertson–Walker models, or more general ones (e.g. Bianchi models, or models without symmetries)? In gravitational terms, will it include only General Relativity, or also brane theories, models with varying G, loop quantum gravity models, string theory models with their associated possibility 'landscapes', and models based on the wave function of the universe concept? Will it allow only standard physics but with varying constants, or a much wider spectrum of physical possibilities, e.g. universes without quantum theory, some with five fundamental forces instead of four, and others with Newtonian gravity? Defining the possibility space means making some kind of assumptions about physics and geometry that will then apply across the whole family of models considered possible in the multiverse, and excluding all other possibilities.

Second, one needs to specify which of the possible universes are physically realized in the multiverse, and how many times each one occurs. *A multiverse must be a physically realized multiverse and not a hypothetical or conceptual one if it is to have genuine explanatory power.* Thus one needs a distribution function $f(m)$ specifying how many times each type of possible universe m in $\mathcal{M}$ is realised. The function $f(m)$ expresses the contingency in any actualization. Things could have been different! Thus, $f(m)$ describes a specific *ensemble of universes* or *multiverse* envisaged as being realised out of the set of possibilities. For example, $f(m)$ might be non-zero for all possible values of all the parameters p ('all that can happen, happens'); but it could be that f describes a multiverse where there are 10^{100} identical copies of one particular universe (the realization process finds a particularly successful recipe, and then endlessly replicates it).

Additionally we need a measure $d\pi$ that enables this function to determine numbers and probabilities of various properties in the multiverse: the number of universes corresponding to a set of parameter increments will be $\mathrm{d}N$ given by

(38) $$\mathrm{d}N = f(m)d\pi$$

for continuous parameters; for discrete parameters, we add in the contribution from all allowed parameter values. The total number of universes N in the ensemble will be given by

(39) $$N = \int_{\mathcal{M}} f(m)d\pi$$

(which will often diverge), where the integral ranges over all allowed values of the member parameters and we take it to include all relevant discrete summations. The expectation value P of a quantity $p(m)$ defined on the set of universes will be given by

(40) $$P = \int_{\mathcal{M}} p(m)f(m)d\pi.$$

These three elements (the possibility space, the measure, and the distribution function) must all be clearly defined in order to give a proper specification of a multiverse [Ellis *et al.*, 2004]. This is almost never done.

9.2.2 Non-uniqueness: Possibilities

There is non-uniqueness at both steps. Stating "all that is possible, happens" does not resolve what is possible. The concept of multiverses is not well defined until the space of possible universes has been fully characterized; it is quite unclear how to do this uniquely. The issue of what is to be regarded as an ensemble of 'all possible' universes can be manipulated to produce any result you want, by redefining what is meant by this phrase — standard physics and logic have no necessary sway over them: what I envisage as 'possible' in such an ensemble may be denied by you. What super-ordinate principles are in operation to control the possibilities in the multiverse, and why? A key point here is that *our understandings of the possibilities are always of necessity arrived at by extrapolation from what we know,* and my imagination may be more fertile than yours, and neither need correspond to what really exists out there — if indeed there is anything there at all. Do we include only

- *Weak variation*: e.g. only the values of the constants of physics are allowed to vary? This is an interesting exercise but is certainly not an implementation of the idea 'all that can happen, happens'. It is an extremely constrained set of variations.

- *Moderate variation*: different symmetry groups, or numbers of dimensions, etc. We might for example consider the possibility landscapes of string theory [Freivogel *et al.*, 2005b] as realistic indications of what may rule multiverses [Susskind, 2003; Freivogel *et al.*, 2005a; Susskind, 2005]. But that is very far indeed from 'all that is possible', for that should certainly include spacetimes not ruled by string theory.

- *Strong variation*: different numbers and kinds of forces, universes without quantum theory or in which relativity is untrue (e.g. there is an aether), some in which string theory is a good theory for quantum gravity and others where it is not, some with quite different bases for the laws of physics (e.g. no variational principles).

- *Extreme variation*: universes where physics is not well described by mathematics; with different logic; universes ruled by local deities; allowing magic as in the Harry Potter series of books; with no laws of physics at all? Without even mathematics or logic?

Which is claimed to be the properties of the multiverse, and why? We can express our dilemma here through the paradoxical question: *Are the laws of logic necessary in all possible universes?*

9.2.3 Non-uniqueness: existence and causation

A specific multiverse is defined by specifying the distribution function $f(m)$ of actually realized universes. It is unclear what mechanism can underlie such a distribution, and any proposal for such a mechanism is completely untestable. We need some indication as to *what determines existence within the possibilities defined by the supposed possibility space*: What decides how many times each one happens?

Unless we understood the supposed underlying mechanisms we can give no serious answer; and there is no prospect whatever of testing any proposed mechanism. The mechanisms supposed to underlie whatever regularities there are in the multiverse must pre-exist the existence of not merely this universe but also every other one. If one assumes a universe that is connected in the large but is locally separated into causally disconnected domains with different physical properties(as in chaotic inflation), one attains a plausible picture of a creation mechanism that can underlie an effective multiverse — but at the expense of supposing the validity of untested and perhaps untestable physics. Because of this one does not obtain a specification of a unique multiverse: the physics could be different than what we assumed.

9.2.4 Explanatory power

What explanatory power do we get in return for these problems? It has been suggested they explain the parameters of physics and of cosmology and in particular the very problematic observed value of the cosmological constant [Weinberg, 2000a; Weinberg, 2000b; Susskind, 2005]. The argument goes as follows: assume a multiverse exists; observers can only exist in one of the highly improbable biophilic outliers where the value of the cosmological constant is very small [Hartle, 2004]. A similar argument has been proposed for neutrino masses [Tegmark *et al.*, 2003]. If the multiverse has many varied locations with differing properties, that may indeed help us understand the Anthropic issue: some regions will allow life to exist, others will not [Barrow and Tipler, 1984; Leslie, 1989]. This does provide a useful modicum of explanatory power. However it is far from conclusive. Firstly, it is unclear why the multiverse should have the restricted kinds of variations of the cosmological constant assumed in the various analyses mentioned. If we assume 'all that can happen, happens' the variations will not be of that restricted kind; those analyses will not apply.

Secondly, ultimate issues remain: Why does this unique larger whole have the properties it does? *Why this multiverse rather than any other one*? Why is it a multiverse that allows life to exist? Many multiverses will not allow any life at all. To solve this, we can propose an *ensemble of ensembles of universes*, with even greater explanatory power and even less prospect of observational verification; and so on. The prospect of an infinite regress looms. Indeed if we declare (as suggested at the start of this article) that 'the Universe' is the total of all that physically exists, then when an ensemble of expanding universe domains exists, whether causally connected or not, that ensemble itself should be called 'the Universe', for it is then the totality of physically existing entities. All the foundational problems for a single existing universe domain recur for the multiverse — because when properly considered, it is indeed the Universe!

9.2.5 Testability

If an ensemble exists with members not connected in any physical way to the observable universe, then we cannot interact with them in any way nor observe them, so we can say anything we like about them without fear of disproof.[32] Thus any statements we make about them can have no solid scientific or explanatory status; they are totally vulnerable to anyone else who claims an ensemble with different properties (for example claiming different kinds of underlying logics are possible in their ensemble, or claiming many physically effective gods and devils in many universes in their ensemble).

Thesis H1: Multiverse proposals are unprovable by observation or experiment, but some self-consistency tests are possible. *Direct observations cannot prove or disprove that a multiverse exists, for the necessary causal relations allowing observation or testing of their existence are absent. Their existence cannot be predicted from known physics, because the supposed causal or pre-causal processes are either unproven or indeed untestable. However some self-consistency conditions for specific multiverse models can be tested.*

Any proposed physics underlying a multiverse proposal, such as Coleman-de Luccia tunneling [Coleman and de Luccia, 1980], will be an extrapolation of known physics; but the validity of that major extrapolation to cosmology is untestable.

Attempts have been made to justify the existence of multiverses as testable firstly via Rees' 'slippery slope' argument [Rees, 2003]. This runs as follows: we can reasonably assume galaxies that we cannot see exist outside the visual horizon (Sec. 8.3); why not extend this argument by small steps to show totally disconnected universes exist? The problem is that this assumes a continuity of existence that does not hold good. The domain outside our horizon is assumed to exist with similar properties to those inside because they are a continuous extension of it and have a largely common causal origin; their nature can be inferred from what we can see. Disconnected multiverse domains are assumed to have quite different properties, and their nature cannot be inferred from what we can see as there is no continuity or causal connection.

Secondly, several authors (Leslie [Leslie, 1989], Weinberg [Weinberg, 2000a; Weinberg, 2000b], and Rees [Rees, 2003] for example) have used arguments based on the idea that the universe is no *more* special than it has to be; a form of "speciality argument." According to Rees, if our universe turns out to be *even more specially* tuned than our presence requires, the existence of a multiverse to explain such "over-tuning" would be refuted; but the actual universe is not more special than this, so the multiverse is not refuted.

In more detail: naive quantum physics predicts the cosmological constant Λ to be very large. But our presence in the universe requires it to be small enough that galaxies and stars can form, so Λ must obviously be below that galaxy-forming threshold. If our universe belongs to an ensemble in which Λ was equally likely to take any value in the biophilic region (the uniform probability assumption),[33] then

[32]But there are counter arguments by Leibniz [Wilson, 1989] and Lewis [Lewis, 1986, section 2.4, pp. 108–115].

[33]The probability distribution for Λ will plausibly peak far away from the biophilic region, tailing down to a low value that will be approximately constant in that narrow region, cf.[Hartle, 2004].

we would not expect it to be too far below this threshold. This is because, if it's too far below the threshold, the probability of randomly choosing that universe in the ensemble becomes very small — there are very few universes with such small values of Λ in the biophilic subset of the ensemble. That is, it would be more likely that any bio-friendly universe in the ensemble would have a value of Λ closer to the threshold value. Present data on this value indicates that it is not too far below the threshold. Thus, our universe is not markedly more special that it needs to be as far as Λ is concerned, and so explaining its fine-tuning by existence of a multiverse is legitimate.

Is this argument compelling? It is a reasonable test of consistency for a multiverse that is known to exist, so that probability considerations apply; but they do not apply if there is no multiverse (Sec. 3). Additionally, probability considerations cannot ever be *conclusive*. Indeed,

Thesis H2: Probability-based arguments cannot demonstrate the existence of multiverses. *Probability arguments cannot be used to* prove *the existence of a multiverse, for they are only applicable if a multiverse exists. Furthermore probability arguments can never prove anything for certain, as it is not possible to violate any probability predictions, and this is* a fortiori *so when there is only one case to consider, so that no statistical observations are possible.*

All one can say on the basis of probability arguments is that some specific state is very improbable. But this does not prove it is impossible, indeed if it is stated to have a low probability, that is precisely a statement that it is possible. Thus such arguments can at best only give plausibility indications even when they are applicable. The assumption that probability arguments can be conclusive is equivalent to the claim that the universe is generic rather than special; but whether this is so or not is precisely the issue under debate (see Thesis **D3**). The argument is useful as a plausibility argument for a multiverse, but is not *proof* of its existence.

Finally, it has been proposed that the existence of multiverses is an inevitable consequence of the universe having infinite space sections [Tegmark, 2003; Seife, 2004], because that leads to infinite spatial repetition of conditions (cf. [Ellis and Brundrit, 1979]). But this supposed spatial infinity is an untested philosophical assumption, which certainly cannot be observationally proven to be correct. Apart from the existence of horizons preventing confirmation of this supposition, even if the entire universe were observable, proving it correct would still not be possible because by definition counting an infinite number of objects takes an infinite amount of time. This is an untestable philosophical argument, not an empirically testable one; furthermore, it can be argued to be implausible (Sec. 9.3.2). Indeed current data suggest it is not the case; this is the one good consistency test one can use for some multiverse proposals (Sec. 9.2.7).

9.2.6 Explanation vs Testability

The argument that this infinite ensemble actually exists can be claimed to have a certain explanatory economy, although others would claim that Occam's razor has been completely abandoned in favour of a profligate excess of existential multiplicity, extravagantly hypothesized in order to explain the one universe that we do know exists. Certainly the price is a lack of testability through either observations or experiment — which is usually taken to be an essential element of any serious scientific theory.[34] It is not uniquely definable nor determinable, and there is a complete loss of verifiability. There is no way to determine the properties of any other universe in the multiverse if they do indeed exist, for they are forever outside observational reach. The point is that there is not just an issue of showing a multiverse exists. If this is a scientific proposition one needs to be able to show which specific multiverse exists; but there is no observational way to do this. Indeed if you can't show *which particular* one exists, it is doubtful you have shown *any* one exists.

What does a claim for such existence mean in this context? Gardner puts it this way: "There is not the slightest shred of reliable evidence that there is any universe other than the one we are in. No multiverse theory has so far provided a prediction that can be tested. As far as we can tell, universes are not even as plentiful as even *two* blackberries" [Gardner, 2003].[35]

Thesis H3: Multiverses are a philosophical rather than scientific proposal. *The idea of a multiverse provides a possible route for the explanation of fine tuning. But it is not uniquely defined, is not scientifically testable apart from some possible consistency tests, and in the end simply postpones the ultimate metaphysical questions.*

The definitive consistency tests on some multiverse proposals (Sec. 9.2.7) are *necessary* conditions for those specific multiverse proposals, but are hardly by themselves indications that the multiverse proposal is true. The drive to believe this is the case comes from theoretical and philosophical considerations (see e.g. [Susskind, 2005]) rather than from data. The claim an ensemble physically exists[36] is problematic as a proposal for scientific explanation, if science is taken to involve testability. Indeed, adopting these explanations is a triumph of theory over testability [Gardner, 2003], but the theories being assumed are not testable. It is therefore a metaphysical choice made for philosophical reasons. That does not mean it is unreasonable (it can be supported by quite persuasive plausibility arguments); but its lack of scientific status should be made clear.

---

[34] In [Stoeger *et al.*, 2004], the framework and conditions under which the multiverse hypothesis would be testable within a retroductive framework, given the rigorous conditions formulated in that paper; are indicated; these conditions are not fulfilled.

[35] This contrasts strongly, for example, with Deutsch's and Lewis's defence of the concept [Deutsch, 1998; Lewis, 1986]. Lewis defends the thesis of "modal realism": that the world we are part of is but one of a plurality of worlds.

[36] As opposed to consideration of an explicitly hypothetical such ensemble, which can indeed be useful, see Sec. 7.2.

9.2.7 Observations and disproof

Despite the gloomy prognosis given above, there are some specific cases where the existence of a chaotic inflation (multi-domain) type scenario (Sec. 2.6) can be *dis*proved. These are firstly when we live in a 'small universe' where we have already seen right round the universe (Sec. 4.3.1), for then the universe closes up on itself in a single FL-like domain, so that no further such domains can exist that are causally connected to us in a single connected spacetime. This 'small universe' situation is observationally testable (Sec. 4.3.1); its confirmation would disprove the usual chaotic inflationary scenario, but not a truly 'disconnected' multiverse proposal, for that cannot be shown to be false by any observation. Neither can it be shown to be true. Secondly, many versions of chaotic inflation, for example those involving Coleman-de Luccia tunneling [Coleman and de Luccia, 1980] from a de Sitter spacetime, demand $k = -1 \Leftrightarrow \Omega_0 < 1$ [Freivogel *et al.*, 2005b; Susskind, 2005]. This requirement is currently marginally disproved by the $2 - \sigma$ bounds on Ω_0 when WMAP observations are combined with the best other available data (Sec. 2.3.7). The best current data is marginally consistent with $k = -1$, but the value indicated most strongly by that data is $k = +1$, indicating finite closed space sections rather than an infinite multiverse such as that advocated by Susskind *et al* [Freivogel *et al.*, 2005b; Susskind, 2005].

9.2.8 Physical or biological paradigms — Adaptive Evolution?

Given that the multiverse idea must in the end be justified philosophically rather than by scientific testing, is there a philosophically preferable version of the idea? One can suggest there is: greater explanatory power is potentially available by introducing the major constructive principle of biology into cosmology, namely adaptive evolution, which is the most powerful process known that can produce ordered structure where none pre-existed. This is realized in principle in Lee Smolin's idea (Sec. 9.1.6) of Darwinian adaptation when collapse to black holes is followed by re-expansion, but with an alteration of the constants of physics each time, so as to allow for evolutionary selection towards those regions that produce the maximum number of black holes. The idea needs development, but is very intriguing:

Thesis H4: The underlying physics paradigm of cosmology could be extended to include biological insights. *The dominant paradigm in cosmology is that of theoretical physics. It may be that it will attain deeper explanatory power by embracing biological insights, and specifically that of Darwinian evolution. The Smolin proposal for evolution of populations of expanding universe domains [Smolin, 1992] is an example of this kind of thinking.*

The result is different in important ways from standard cosmological theory precisely because it embodies in one theory three of the major ideas of last century, namely (i) Darwinian evolution of populations through competitive selection, (ii) the evolution of the universe in the sense of major changes in its structure associated with its expansion, and (iii) quantum theory, underlying the only partly explicated mechanism supposed to cause re-expansion out of collapse into a black hole. There is a great contrast with the theoretical physics paradigm of dynamics governed simply by variational principles shaped by symmetry considerations. It seems worth pursuing as a very different route to the understanding of the creation of structure.[37]

9.3 Issue I: Natures of Existence

Underlying all this is the issue of natures of existence, which has a number of aspects, relating from the purely physical to more metaphysical issues.

9.3.1 Physical existence: kinds of matter

Unsolved key issues for physical cosmology relate to what kind of matter and/or fields exist. While we understand matter in the solar system quite well, at present we do not understand most of what exists in the universe at large:

Thesis I1: We do not understand the dominant dynamical matter components of the universe at early or late times. *A key goal for physical cosmology is determining the nature of the inflaton, of dark matter, and of dark energy. Until this is done, the causal understanding of cosmology is incomplete, and in particular the far future fate of the universe is unknown.*

This is the core activity of much work in cosmology at present. Until they are all explicated, cosmology is not properly linked to physics, and the nature of the matter that dominates the dynamics of the universe is unknown. Its explication is surely one of the key concerns of cosmology [Durrer, 2002]. A key requirement is that even if we cannot experimentally verify the proposed nature of the matter, at least it should be physically plausible. This appears not to be the case for some current proposals, e.g. so-called 'phantom matter' which has negative kinetic energy terms.

The far future fate of the universe depends crucially on the effective equation of state for dark matter ('quintessence'). But the problem is that even if we can determine these properties at the present time (for one particular range of parameter values), this does not necessarily guarantee what they will be in the far future (for a quite different range of parameter values that are probably outside the range of possible experimental test). Furthermore adjusting a 'dark energy' model to fit the supernova data does not determine the underlying physics. One can fit any monotonic evolution $S(t)$ with a suitable choice of the equation of state function $p(\mu)$. Specifically, for any $S(t)$ and any k we define $\mu(t)$ and $p(t)$ by

$$(41)\quad \kappa\mu(t) = 3\left[\frac{\dot{S}^2(t)}{S^2(t)} + \frac{k}{S^2(t)}\right],\ \kappa p(t) = \left[\frac{\dot{S}^2(t)}{S^2(t)} + \frac{k}{S^2(t)}\right] - 2\frac{\ddot{S}(t)}{S(t)},$$

[37]Cf. Chapter 13 of Susskind [Susskind, 2005].

then (9), (7) will be exactly satisfied, and we have 'solved' the field equations for this arbitrarily chosen monotonic evolution $S(t)$. If we can observationally determine the form of $S(t)$, for example from (m, z)–curves associated with supernovae data, this is essentially how we can then determine that some kind of 'dark energy' or 'quintessence' is required to give that evolution, and we can find the equation of state implied by eliminating t between these two equations. This is, however, not a *physical* explanation until we have either in some independent experimental test demonstrated that matter of this form exists, or have theoretically shown why this matter or field has the form it does in some more fundamental terms than simply a phenomenological fit. If we assume the matter is a scalar field, the kinetic energy term $\dot{\phi}^2$ implied by (32), (41) may be negative — which is the case for so-called 'shadow matter' models proposed recently by some worker. If normal physics criteria are applied, this is a proof that this kind of matter is unphysical, rather than an identification of the nature of the dark energy.

9.3.2 Existence of Infinities

The nature of existence is significantly different if there is a finite amount of matter or objects in the universe, as opposed to there being an infinite quantity in existence. Some proposals claim there may be an infinite number of universes in a multiverse and many cosmological models have spatial sections that are infinite, implying an infinite number of particles, stars, and galaxies. However, infinity is quite different from a very large number! Following David Hilbert [Hilbert, 1964], one can suggest these unverifiable proposals cannot be true: the word 'infinity' denotes a quantity or number that can never be attained, and so will never occur in physical reality.[38] He states

"Our principal result is that the infinite is nowhere to be found in reality. It neither exists in nature nor provides a legitimate basis for rational thought . . . The role that remains for the infinite to play is solely that of an idea . . . which transcends all experience and which completes the concrete as a totality . . ." [Hilbert, 1964, p. 151].

This suggests "infinity" cannot be arrived at, or realized, in a concrete physical setting; on the contrary, the concept itself implies its inability to be realized![39]

Thesis I2: The often claimed physical existence of infinities is questionable. *The claimed existence of physically realized infinities in cosmology or multiverses raises problematic issues. One can suggest they are unphysical; in any case such claims are certainly unverifiable.*

This applies in principle to both small and large scales in any single universe:

[38] An intriguing further issue is the dual question: Does the quantity zero occur in physical reality? This is related to the idea of physical existence of nothingness, as contrasted with a vacuum [Seife, 2000]. A vacuum is not nothing! (cf. [Susskind, 2005]).

[39] For a contrasting view, see Bernadete [Bernadete, 1964].

- The existence of a physically existing spacetime continuum represented by a real (number) manifold at the micro-level contrasts with quantum gravity claims of a discrete spacetime structure at the Planck scale, which one might suppose was a generic aspect of fully non-linear quantum gravity theories [Rovelli, 2004]. In terms of physical reality, this promises to get rid of the uncountable infinities the real line continuum engenders in all physical variables and fields.[40] There is no experiment that can *prove* there is a physical continuum in time or space; all we can do is test space-time structure on smaller and smaller scales, but we cannot approach the Planck scale.

- Infinitely large space-sections at the macro-level raise problems as indicated by Hilbert, and leads to the infinite duplication of life and all events [Ellis and Brundrit, 1979]. We may assume space extends forever in Euclidean geometry and in many cosmological models, but we can never prove that any realised 3-space in the real universe continues in this way — it is an untestable concept, and the real spatial geometry of the universe is almost certainly not Euclidean. Thus Euclidean space is an abstraction that is probably not physically real. The infinities supposed in chaotic inflationary models derive from the presumption of pre-existing infinite Euclidean space sections, and there is no reason why those should necessarily exist. In the physical universe spatial infinities can be avoided by compact spatial sections, resulting either from positive spatial curvature, or from a choice of compact topologies in universes that have zero or negative spatial curvature. Machian considerations to do with the boundary conditions for physics suggest this is highly preferable [Wheeler, 1968]; and if one invokes string theory as a fundamental basis for physics, then 'dimensional democracy' suggests the three large spatial dimensions should also be compact, since the small ('compactified') dimensions are all taken to be so. The best current data from CBR and other observations (Sec. 2.3.7) indeed suggest $k = +1$, implying closed space sections for the best-fit FL model.

- The existence of an eternal universe implies that an infinite time actually exists, which has its own problems: if an event happens at any time t_0, one needs an explanation as to why it did not occur before that time (as there was an infinite previous time available for it to occur); and Poincaré eternal return (mentioned in Sec. 6.1) will be possible if the universe is truly cyclic. In any case it is not possible to *prove* that the universe as a whole, or even the part of the universe in which we live, is past infinite; observations cannot do so, and the physics required to guarantee this would happen (if initial conditions were right) is untestable. Even attempting to prove it is future infinite is problematic (we cannot for example guarantee the properties of the

[40] To avoid infinities entirely would require that nothing whatever is a continuum in physical reality (since any continuum interval contains an infinite number of points). Doing without that, conceptually, would mean a complete rewrite of many things. Considering how to do so in a way compatible with observation is in my view a worthwhile project.

vacuum into the infinite future — it might decay into a state corresponding to a negative effective cosmological constant).

- It applies to the possible nature of a multiverse. Specifying the geometry of a generic universe requires an infinite amount of information because the quantities necessary to do so are fields on spacetime, in general requiring specification at each point (or equivalently, an infinite number of Fourier coefficients): they will almost always not be algorithmically compressible. All possible values of all these components in all possible combinations will have to occur in a multiverse in which "all that can happen, does happen". There are also an infinite number of topological possibilities. This greatly aggravates all the problems regarding infinity and the ensemble. Only in highly symmetric cases, like the FL solutions, does this data reduce to a finite number of parameters, each of which would have to occur in all possible values (which themselves are usually taken to span an infinite set, namely the entire real line). Many universes in the ensemble may themselves have infinite spatial extent and contain an infinite amount of matter, with all the problems that entails. To conceive of physical creation of an infinite set of universes (most requiring an infinite amount of information for their prescription, and many of which will themselves be spatially infinite) is at least an order of magnitude more difficult than specifying an existent infinitude of finitely specifiable objects.

One should note here particularly that problems arise in the multiverse context from the continuum of values assigned by classical theories to physical quantities. Suppose for example that we identify corresponding times in the models in an ensemble and then assume that *all* values of the density parameter and the cosmological constant occur at each spatial point at that time. Because these values lie in the real number continuum, this is a doubly uncountably infinite set of models. Assuming genuine physical existence of such an uncountable infinitude of universes is the antithesis of Occam's razor. But on the other hand, if the set of realised models is either finite or countably infinite, then almost all possible models are not realised. And in any case this assumption is absurdly unprovable. We can't observationally demonstrate a single other universe exists [Gardner, 2003], let alone an infinitude. The concept of infinity is used with gay abandon in some multiverse discussions [Knobe *et al.*, 2005], without any concern either for the philosophical problems associated with this statement [Hilbert, 1964], or for its completely unverifiable character. It is an extravagant claim that should be treated with extreme caution.

9.3.3 The Nature of the Laws of Physics

Underlying all the above discussion is the basic concept of ordered behaviour of matter, characterized by laws of physics of a mathematical nature that are the same everywhere in the universe.[41] Three interlinked issues arise.

(i) *What is the ontological nature of the laws of physics*: descriptive, *just characterizing the way things are, or* prescriptive, *enforcing them to be this way*? [Carroll, 2004]. If they are descriptive, the issue arising is, *Why does all matter have the same properties wherever it exists in the universe*? Why are all electrons everywhere in the universe identical, if the laws are only descriptive? If they are prescriptive, then matter will necessarily be the same everywhere (assuming the laws themselves are invariable); the issue arising then is, *In what way do laws of physics exist that enforce themselves on the matter in the universe*? Do they for example have an existence in some kind of Platonic space that controls the nature of matter and existence? One can avoid talking about the laws of physics *per se* by instead considering the *space of possibilities* underlying what exists physically, rigorously constraining the possible natures of what actually comes into existence [Ellis, 2004]. This space is more or less uniquely related to the underlying laws in the same way that the space of solutions of differential equations is related to the nature of the equations. This enables one to avoid the issue of the ontology of the laws of physics, but does not solve it.

(ii) *Why are the laws of physics so well explained by mathematical descriptions*? If they are prescriptive, this deep issue might be related to the suggested Platonic nature of the space of mathematical reality [Penrose, 2004]. If they are descriptive, then the mathematical expressions we use to encapsulate them are just a convenient description but do not reflect their ultimate nature. Many writings in physics and cosmology seem to assume that their ultimate existential nature is indeed mathematical — perhaps a confusion of appearance and reality (see Sec. 8.3).

(iii) *Do they pre-exist the universe and control its coming into being, or do they come into being with the universe*? This is where this issue relates deeply to the nature of cosmology, and is clearly related to the other two questions raised above. Many theories of creation of the universe assume that all these laws, or at least a basic subset, pre-exist the coming into being of the physical universe, because they are presumed to underlie the creation process, for example the entire apparatus of quantum field theory is often taken for granted as pre-existing our universe (Sec. 6). This is of course an unprovable proposition

Thesis I3: A deep issue underlying the nature of cosmology is the nature of the laws of physics. *The nature of the possibility space for physical existence is characterized by the laws of physics. However it is unclear if these laws are prescriptive or descriptive; whether they come into being with space-time and matter, or pre-exist them.*

[41] The effective laws may vary from place to place because for example the vacuum state varies [Susskind, 2005]; but the fundamental laws that underlie this behaviour are themselves taken to be invariant.

9.3.4 'Ultimate Reality'

Philosophers have debated for millennia whether the ultimate nature of existence is purely material, or embodies some form of rationality ('Logos') and/or purpose ('Telos'). *What in the end underlies it all*? Is the ultimate nature of the universe purely material, or does it in some way have an element of the mental? (cf. Sec. 9.1.6). That profound debate is informed by physical cosmology, but cannot be resolved by the physical sciences alone (Sec. 9.1.7). Here, I will make just two comments on this deep issue.

Firstly, even in order to understand just the material world, it can be claimed that one needs to consider forms of existence other than the material only — for example a Platonic world of mathematics and a mental world, both of which can be claimed to exist and be causally effective in terms of affecting the material world [Ellis, 2004; Penrose, 2004]. Our understanding of local causation will be incomplete unless we take them into account.

Secondly, in examining these issues one needs to take into account data about the natures of our existence that come from our daily lives and the broad historical experience of humanity (our experiences of ethics and aesthetics, for example), as well as those discoveries attained by the scientific method. Many writings claim there is no purpose in the universe: it is all just a conglomerate of particles proceeding at a fundamental level in a purposeless and meaningless algorithmic way. But I would reply, the very fact that those writers engage in such discourse undermines their own contention; they ignore the evidence provided by their own actions. There is certainly meaning in the universe to this degree: *the fact they take the trouble to write such contentions is proof that they consider it meaningful to argue about such issues*; and this quality of existence has emerged out of the nature of the physical universe (Sec. 7.3). Indeed the human mind is causally effective in the real physical world precisely through many activities motivated by meanings perceived by the human mind. Any attempt to relate physics and cosmology to ultimate issues must take such real world experience seriously [Ellis, 2005], otherwise it will simply be ignoring a large body of undeniable data. This data does not resolve the ultimate issues, but does indicate dimensions of existence that indeed do occur.

10 CONCLUSION

The physical scale of the Universe is enormous, and the images of distant objects from which we obtain our information are extremely faint. It is remarkable that we are able to understand the Universe as well as we do. An intriguing feature is the way in which the philosophy of cosmology is to a considerable degree shaped by contingent aspects of the nature of the universe — its vast scale (Sec. 4), leading to the existence of visual horizons (Sec. 4.3), and the occurrence of extreme energies in the early universe (Sec. 5), leading to the existence of physical horizons. Philosophical issues arising in relation to cosmology (Sec. 8) would be quite different if its physical structure were very different. Furthermore in order that philosophical analysis can engage with cosmology in depth, the detailed nature of the relation between observations and theory in cosmology (Sec. 2) is relevant.

10.1 Are there laws of cosmology?

As we have discussed in detail, the uniqueness of the universe implies the unique nature of cosmology. We now return to the initial issue, *Are there Laws of the Universe*? (Sec. 3). At one level, the laws of the cosmos are simply the local laws we know and love (e.g. Maxwell's laws, Einstein's field equations) applied to the whole shebang. Of course, there is the problem of extrapolation from the local to the global. But although the extrapolation is bigger in cosmology, it seems not to be different in kind from what we always do in science. In that sense, there are no special laws for the evolution of the universe. But that does not determine the outcome: cosmology needs some prescription of boundary or initial conditions as well, in order to determine the future. Is there a true "Cosmological principle", a law of initial conditions for the universe, that determines this outcome?

The idea of "Laws of initial conditions of the universe" seems not to be a testable idea (Sec. 3). Scientifically, one can only describe what occurred rather than relate it to generic principles, for such principles cannot be tested. In fact any description of boundary or initial conditions for the universe seems to be just that: a description of these conditions, rather than a testable prescription of how they must be. The 'Cosmological Principle' — the universe is necessarily spatially homogeneous and isotropic (Sec. 4.2.2) is of this kind: a description of the way the initial data turned out, rather than a fundamental reason for why this should be so. Justification of this view was based by some workers on a *Copernican Principle* (the assumption we do not live in a privileged place in the universe), strengthened to become a *Cosmological Principle* [Bondi, 1960; Weinberg, 1972; Harrison, 2000]; but this is a philosophical assumption — essentially, a claim that the universe must have very special initial conditions — which may or may not be true, and does not attempt a physical explanation. This kind of argument is out of fashion at present, because we now prefer generality to speciality and physical argumentation to geometrical prescription; but it was previously strongly proposed (e.g. [Weinberg, 1972], pp. 407-412). The tenor of philosophical argument has changed.

Nevertheless there is one kind of Law of the Universe one might propose, following McCrea [McCrea, 1970]: namely an "Uncertainty principle in cosmology", dual to the uncertainty principle in quantum theory. Uncertainty applies on the largest scale, as we have discussed above in some detail, and also on the smallest, where it is a profound feature of quantum theory. Its basis is very different in the two cases, on the one hand (in quantum theory) being ontological in nature, on the other (in cosmology) being epistemological in nature.[42] Nevertheless it is

[42] Assuming that quantum uncertainty is indeed ontological rather than epistemological. One should however keep an open mind on this: just because it is the current dogma does not

a key aspect of our relation to the cosmos, so that (following McCrea) we might perhaps formalize it in order to emphasize its centrality to the relation between cosmology and philosophy:

Thesis of Uncertainty: Ultimate uncertainty is a key aspect of cosmology. *Scientific exploration can tell us much about the universe but not about its ultimate nature, or even much about some of its major geometrical and physical characteristics. Some of this uncertainty may be resolved, but much will remain. Cosmological theory should acknowledge this uncertainty.*

10.2 What can we truly claim

Cosmology considers questions of physical origins in the uniquely existing physical universe (Sec. 6) which provides the context of our existence (Sec. 7, Sec. 9.1). These questions can be extended to include ultimate issues if we so desire (Sec. 8.2), but physical theory cannot resolve them (Sec. 9.1.7). In the end, there are a variety of mysteries underlying the existence and nature of the universe (Sec. 9.3). The scientific study of cosmology can help illuminate their nature, but cannot resolve them.

As well as celebrating the achievements of cosmology, one should fully take into account the limits and problems considered in this chapter, and not claim for scientific cosmology more than it can actually achieve or more certainty than is in fact attainable. Such claims will in the long term undermine cosmology's legitimate status as a project with solid scientific achievements to its name. That status can be vigorously defended as regards the 'Standard Model' of cosmology (Sec. 2.8), provided this standard model is characterized in conservative terms so that it is not threatened by relatively detailed shifts in theory or data that do not in fact threaten the core business of cosmology. Further, this defence must take adequate cognisance of the difficult philosophical issues that arise if one pushes the explanatory role of cosmological theory to its limits (Sec. 6); for example one should not make too strong *scientific* claims in regard to the possible existence of multiverses (Sec. 9.2); philosophically based plausibility arguments for them are fine, if identified as such. Cosmology is not well served by claims that it can achieve more explanatory power than is in fact attainable, or by statements that its claims are verified when in fact the requisite evidence is unavailable, and in some cases must forever remain so.

ACKNOWLEDGEMENT

I thank Bill Stoeger, Martin Bojowald, Malcolm MacCallum, Henk van Elst, Jeremy Butterfield, and John Earman for useful comments that have improved this article.

necessarily mean it is true.

Abbreviations used:

CBR: Cosmic Blackbody Radiation, p. 1192
CDM: Cold Dark Matter, p. 1201
EFE: Einstein Field Equations, p. 1185
FL: Friedmann-Lemaître (universe models), p. 1187
HBB: Hot Big Bang, p. 1191
LSS: Last Scattering surface, p. 1192
RW: Robertson-Walker (geometry), p. 1186
SAP: Strong Anthropic Principle, p. 1254
WAP: Weak Anthropic Principle, p. 1253

BIBLIOGRAPHY

[Allday, 2002] J. Allday. *Quarks, Leptons and the Big Bang.* Institute of Physics, 2002.

[Balashov, 1991] Y. V. Balashov. Resource letter Ap-1: The anthropic principle. *American Journal of Physics*, 54:1069–1076, 1991.

[Barbour and Pfister, 1995] J. Barbour and H. Pfister. *Mach's Principle: From Newton's Bucket to Quantum Gravity.* Birkhäuser, Basel, 1995.

[Bardeen, 1980] J. M. Bardeen. Gauge-invariant cosmological perturbations. *Phys. Rev.*, D22:1882-1905, 1980.

[Barnett *et al.*, 1996] R. M. Barnett *et al.*(Particle Data Group) Particle physics summary. a digest of the 1996 Review of Particle Physics. *Reviews of Modern Physics*, 68:611–732, 1996.

[Barrow and Tipler, 1984] J. Barrow and F. Tipler. *The Cosmological Anthropic Principle.* Oxford University Press, 1984.

[Barrow, 2003] J. D. Barrow. *The Constants of Nature.* Vintage, 2003.

[Bennet et al., 2003] C. L. Bennet *et al.* (the WMAP team). First year Wilkinson Microwave Anisotropy Probe (WMAP) observations: Preliminary maps and basic results. *Astrophysical Journal Supplement Series*, 148:1–27, 2003. astro-ph/0302207.

[Berendzen *et al.*, 1976] R. Berendzen, R. Hart and D. Seeley. *Man Discovers the Galaxies.* Science History Publications, New York, 1976.

[Bernadete, 1964] J. A. Bernadete. *Infinity, An Essay in Metaphysics.* Clarendon Press, Oxford, 1964.

[Blau and Guth, 1987] S. K. Blau and A. H. Guth. Inflationary cosmology. In Hawking, S. W. and Israel, W., editors, *300 Years of Gravitation*, page 524. Cambridge University Press, 1987.

[Bojowald, 2001] M. Bojowald. Absence of singularity in loop quantum cosmology. *Physical Review Letters*, 86:5227–5239, 2001.

[Bondi, 1947] H. Bondi. Spherically symmetric models in general relativity. *Monthly Notice of the Royal Astronomical Society*, 107:410, 1947.

[Bondi, 1960] H. Bondi. *Cosmology.* Cambridge University Press, Cambridge, 1960.

[Bonnor and Ellis, 1986] W. B. Bonnor and G. F. R. Ellis. Observational homogeneity of the universe. *Monthly Notice of the Royal Astronomical Society*, 218:605–614, 1986.

[Bostrom, 2002] N. Bostrom. *Anthropic bias: Observation selection effects in science and philosophy.* Routledge, New York, 2002.

[Bothun, 1998] G. Bothun. *Modern Cosmological Observations and Problems.* Taylor and Francis, 1998.

[Brading and Castellani, 2006] K. Brading and E. Castellani. This vol, Ch. 13, 2006.

[Brans and Dicke, 1961] C. Brans and R. H. Dicke. Mach's principle and a relativistic theory of gravitation. *Physical Review*, 124:925, 1961.

[Carr and Rees, 1979] B. J. Carr and M. J. Rees. The anthropic principle and the structure of the physical world. *Nature*, 278:605–612, 1979.

[Carroll, 2004] J. Carroll, editor. *Readings on Laws of Nature.* University of Pittsburgh Press, 2004.

[Carroll and Chen, 2005] S. M. Carroll and J. Chen. Does inflation provide natural initial conditions for the universe? gr-qc/0505037, 2005.

[Challinor and Lasenby, 1998] A. D. Challinor and A. N. Lasenby. Cosmic microwave background anisotropies in the CDM model: A covariant and gauge-invariant approach. *Physical Review D*, 58. astro-ph/9804301, 1998.

[Coleman and de Luccia, 1980] S. Coleman and F. de Luccia. Gravitational effects on and of vacuum decay. *Physical Review D*, 21:3305–3315, 1980.

[Copeland *et al.*, 2005] E. J. Copeland, J. E. Lidsey, and S. Mizuno. Correspondence between loop-inspired and braneworld cosmology. gr-qc/0510022, 2005.

[Cowie and Songaila, 1995] L. Cowie and A. Songaila. Astrophysical limits on the evolution of dimensionless physical constants over cosmological time. *Astrophysics Journal*, 453:596–598, 1995.

[Davies, 1974] P. C. W. Davies. *The physics of time asymmetry.* Surrey University Press, London, 1974.

[Davies, 1982] P. C. W. Davies. *The Accidental Universe.* Cambridge University Press, 1982.

[Davies, 1987] P. C. W. Davies. *The cosmic Blueprint.* Heinemann, London, 1987.

[Davis and Lineweaver, 2004] T. M. Davis. and C. H. Lineweaver. Expanding confusion: common misconceptions of cosmological horizons and the superluminal expansion of the universe. *Publications of the Astronomical Society of Australia*, 21:97–109, 2004. astro-ph/0310808.

[de Vaucouleurs, 1970] G. de Vaucouleurs. The case for a hierarchical cosmology. *Science*, 167:1203–13, 1970.

[Deutsch, 1998] D. Deutsch. *The Fabric of reality: The science of parallel universes - and its implications.* Penguin, 1998.

[Dicke and Peebles, 1979] R. H. Dicke. and P. J. E. Peebles. The big bang cosmology—enigmas and nostrums. In Hawking, S. W. and Israel, W., editors, *General Relativity: An Einstein Centenary Survey*, pages 504–517. Cambridge University Press, Cambridge, 1979.

[Dickson, 2006] M. I. Dickson. this vol , Ch.4, 2006.

[Dirac, 1938] P. A. M. Dirac. New basis for cosmology. *Proceedings of the Royal Society of LondoAn*, 165:199–208, 1938.

[Disney, 1976] M. J. Disney. Visibility of galaxies. *Nature*, 263:573–575, 1976.

[Dodelson, 2003] S. Dodelson. *Modern Cosmology.* Academic Press, 2003.

[Durrer, 2002] R. Durrer. Frontiers of the universe: What do we know and what do we understand? astro-ph/0205101, 2002.

[Earman, 1987] J. Earman. The SAP also rises: A critical examination of the anthropic principle. *American Philosophical Quarterly*, 24:307–325, 1987.

[Earman, 1999] J. Earman. The Penrose-Hawking singularity theorems: History and implications. In Goenner, H., Renn, J., and Sauer, T., editors, *The Expanding Worlds of General Relativity.* Birkhäuser, 1999.

[Earman, 2001] J. Earman. Lambda: The constant that would not die. *Archive for History of Exact Sciences*, 55:189–220, 2001.

[Earman, 2003] J. Earman. The cosmological constant, the fate of the universe, unimodular gravity, and all that. *Studies in History and Philosophy of Modern Physics*, 34:559–577, 2003.

[Earman and Mosterin, 1999] J. Earman and J. Mosterin. A critical look at inflationary cosmology. *Philosophy of Science*, 66:1–49, 1999.

[Ehlers, 1993] J. Ehlers. Contributions to the relativistic mechanics of continuous media. *General Relativity and Gravitation*, 25:1225–1266, 1993.

[Ehlers *et al.*, 1968] J. Ehlers, P. Geren, and R. K. Sachs. Isotropic solutions of the Einstein-Liouville equations. *Journal of Mathematical Physics*, 9:1344–1349, 1968.

[Ehlers and Rindler, 1989] J. Ehlers and W. Rindler. A phase-space representation of Friedmann-Lemaître universes containing both dust and radiation and the inevitability of a big bang. *Monthly Notice of the Royal Astronomical Society*, 238:503–521, 1989.

[Eisenstein *et al.*, 2005a] D. J. Eisenstein *et al.* Detection of the baryon accoustic peak in the large scale correlation function of galaxies, 2005. astro-ph/0501171.

[Ellis, 1967] G. F. R. Ellis. The dynamics of pressure-free matter in general relativity. *Journal of Mathematical Physics*, 8:1171–1194, 1967.

[Ellis, 1971a] G. F. R. Ellis. Relativistic cosmology. In Sachs, R. K., editor, *General Relativity and Cosmology, Proceedings of the XLVII Enrico Fermi Summer School.* Academic Press, New York, 1971.

[Ellis, 1971b] G. F. R. Ellis. Topology and cosmology. *General Relativity and Gravitation*, 2:7–21, 1971.

[Ellis, 1975] G. F. R. Ellis. Cosmology and verifiability. *Quarterly Journal of the Royal Astronomical Society*, 16:245–264, 1975.

[Ellis, 1980] G. F. R. Ellis. Limits to verification in cosmology. *Ann New York Acad Sci*, 336:130–160, 1980.

[Ellis, 1984] G. F. R. Ellis. Relativistic cosmology: its nature, aims and problems. In *et al.*, B. B., editor, *General Relativity and Gravitation*, pages 215–288. Reidel, 1984.

[Ellis, 1989] G. F. R. Ellis. A history of cosmology 1917-1955. In Howard, D. and Stachel, J., editors, *Einstein and the History of General Relativity*, volume 1 of *Einstein Study Series*, pages 367–431. Birkhäuser, Boston, 1989.

[Ellis, 1990] G. F. R. Ellis. Innovation resistance and change: the transition to the expanding universe. In Bertotti, B., Balbinto, R., Bergia, S., and Messina., A., editors, *Modern Cosmology in Retrospect*, pages 97–114. Cambridge University Press, 1990.

[Ellis, 1991] G. F. R. Ellis. Major themes in the relation between philosophy and cosmology. *Mem Ital Ast Soc*, 62:553–605, 1991.

[Ellis, 2002] G. F. R. Ellis. Cosmology and local physics. *New Astronomy Reviews*, 46:645–658, 2002. gr-qc/0102017.

[Ellis, 2004] G. F. R. Ellis. True complexity and its associated ontology. In Barrow, J. D., Davies, P. C., , and Jr., C. L. H., editors, *Science and Ultimate Reality: Quantum Theory, Cosmology and Complexity, honoring John Wheeler's 90th birthday.* Cambridge University Press, 2004.

[Ellis, 2005] G. F. R. Ellis. Physics, complexity, and causality. *Nature*, 435:743, 2005.

[Ellis and Baldwin, 1984] G. F. R. Ellis. and J. Baldwin. On the expected anisotropy of radio source counts. *Monthly Notice of the Royal Astronomical Society*, 206:377–381, 1984.

[Ellis and Brundrit, 1979] G. F. R. Ellis and G. B. Brundrit. Life in the infinite universe. *Quarterly Journal of the Royal Astronomical Society*, 20:37–41, 1979.

[Ellis and Bruni, 1989] G. F. R. Ellis and M. Bruni. A covariant and gauge-free approach to density fluctuations in cosmology. *Physical Review D*, 40:1804–1818, 1989.

[Ellis and Buchert, 2006] G. F. R. Ellis and T. R. Buchert. The universe seen at different scales. To appear, *Physics Letter A*, special Einstein issue, 2006. gr-qc/0506106.

[Ellis and King, 1974] G. F. R. Ellis and A. R. King. Was the big bang a whimper? *Communications in Mathematical Physics*, 38:119–165, 1974.

[Ellis *et al.*, 2004] G. F. R. Ellis, U. Kirchner, and W. Stoeger. Multiverses and physical cosmology. *Monthly Notice of the Royal Astronomical Society*, 347:921–936, 2004.

[Ellis and Maartens, 2004] G. F. R. Ellis and R. Maartens. The emergent universe: inflationary cosmology with no singularity and no quantum gravity era. *Classical and Quantum Gravity*, 21:223–232, 2004.

[Ellis *et al.*, 1978] G. F. R. Ellis, R. Maartens, and S. D. Nel. The expansion ofthe universe. *Monthly Notice of the Royal Astronomical Society*, 184:439–465, 1978.

[Ellis and Madsen, 1991] G. F. R. Ellis and M. Madsen. Exact scalar field cosmologies. *Classical and Quantum Gravity*, 8:667–676, 1991.

[Ellis *et al.*, 2002a] G. F. R. Ellis, P. McEwan, W. Stoeger, and P. Dunsby. Causality in inflationary universes with positive spatial curvature. *General Relativity and Gravitation.*, 34:1461–1481, 2002. gr-qc/0109024.

[Ellis *et al.*, 1985] G. F. R. Ellis, S. D. Nel, W. Stoeger, R. Maartens, and A. P. Whitman. Ideal observational cosmology. *Physics Reports*, 124:315–417, 1985.

[Ellis *et al.*, 1984] G. F. R. Ellis, J. J. Perry, and A. Sievers. Cosmological observations of galaxies: the observational map. *Astronomical Journal*, 89:1124–1154, 1984.

[Ellis and Schreiber, 1986] G. F. R. Ellis and G. Schreiber. Observational and dynamic properties of small universes. *Physics Letters A*, 115:97–107, 1986.

[Ellis and Sciama, 1972] G. F. R. Ellis and D. W. Sciama. Global and non-global problems in cosmology. In O'Raifeartaigh, L., editor, *General Relativity*, pages 35–59. Oxford University Press, 1972.

[Ellis *et al.*, 2002b] G. F. R. Ellis, W. Stoeger, P. McEwan, and P. Dunsby. Dynamics of inflationary universes with positive spatial curvature. *General Relativity and Gravitation*, 34:1445–1459, 2002. gr-qc/0109023.

[Ellis and Stoeger, 1987] G. F. R. Ellis and W. R. Stoeger. The Fitting problem in cosmology. *Classical and Quantum Gravity*, 4:1679–1690, 1987.

[Ellis and Stoeger, 1988] G. F. R. Ellis and W. R. Stoeger. Horizons in inflationary universes. *Classical and Quantum Gravity*, 5:207, 1988.

[Ellis and Uzan, 2005] G. F. R. Ellis and J.-P. Uzan. 'c' is the speed of light, isn't it? *American Journal of Physics*, 73:240–247, 2005.

[Ellis and van Elst, 1999a] G. F. R. Ellis and H. van Elst. Cosmological models (cargese lectures 1998). In Lachièze-Ray, M., editor, *Theoretical and Observational Cosmology*, volume 541, pages 1–116. Kluwer, Nato Series C: Mathematical and Physical Sciences, 1999. gr-qc/9812046.

[Ellis and van Elst, 1999b] G. F. R. Ellis and H. van Elst. Deviation of geodesics in FLRW spacetime geometries. In Harvey, A., editor, *On Einstein's Path: Essays in Honour of Englebert Schucking*, pages 203–226. Springer, 1999.

[Ellis and Williams, 2000] G. F. R. Ellis and R. M. Williams. *Flat and Curved Spacetimes.* Cambridge Univesity Press, 2000.

[Fabian, 1989] A. C. Fabian, editor. *Origins*. Cambridge University Press, 1989.

[Freivogel *et al.*, 2005a] B. Freivogel, M. Kleban, M. R. Martinez, and L. Susskind. Observational consequences of a landscape. hep-th/0505232, 2005.

[Freivogel *et al.*, 2005b] B. Freivogel, M. Kleban, M. R. Martinez, and L. Susskind. Observational consequences of a landscape. hep-th/0505232, 2005.

[Gardner, 1986] M. Gardner. WAP, SAP, PAP, and FAP. *The New York Review of Books*, 23(8):22–25, 1986.

[Gardner, 2003] M. Gardner. *Are Universes Thicker than Blackberries?* W W Norton, 2003.

[Garrett and Coles, 1993] A. J. M. Garrett and P. Coles. Bayesian inductive inference and the anthropic cosmological principle. *Comments Astrophys*, 17:23–47, 1993.

[Ghosh *et al.*, 1998] A. Ghosh, G. Pollifrone, and G. Veneziano. Quantum fluctuations in open pre-big bang cosmology. *Physics LetterBs*, 440:20–27, 1998.

[Gibbons *et al.*, 1983] G. W. Gibbons, S. W. Hawking, and S. T. C. Siklos. *The Very Early Universe*. Cambridge University Press, 1983.

[Gibbons *et al.*, 2003] G. W. Gibbons, E. P. S. Shellard, and S. J. Rankin, editors. *The Future of Theoretical Physics and Cosmology: Celebrating Stephen Hawking's 60th Birthday.* Cambridge University Press, 2003.

[Goodman, 1995] J. Goodman. Geocentrism reexamined. *Physical Review D*, 52:1821–1827, 1995.

[Gott and Li, 1998] J. R. Gott and L.-X. Li. Can the universe create itself? *Physical Review D*, 58:023501, 1998. astro-ph/9712344.

[Gribbin and Rees, 1991] J. Gribbin and M. Rees. *Cosmic Coincidences*. Black Swan, 1991.

[Grunbaum, 1989] A. Grunbaum. The pseudo problem of creation. *Philosophy of Science*, 56:373–394, 1989. reprinted in [Leslie, 1998].

[Grunbaum, 2004] A. Grunbaum. The poverty of theistic cosmology. *British Journal of Philosophy of Science*, 55:561–614, 2004.

[Gunn *et al.*, 1978] J. E. Gunn, M. S. Longair and M. J. Rees. *Observational cosmology: Saas-Fee Course 8*. Geneva Observatory, Switzerland, 1978.

[Guth, 1997] A. H. Guth. *The Inflationary Universe: The Quest for a new Theory of Cosmic Origins*. Addison Wesley, 1997.

[Guth, 1981] A. H. Guth. Inflationary universe: A possible solution to the horizon and flatness. *Physical Review D*, 23:347, 1981.

[Guth, 2001] A. H. Guth. Eternal inflation. In *Proceedings of "Cosmic Questions"Meeting*, volume 95. The New York Academy of Sciences Press, 2001. astro-ph/0101507.

[Harris *et al.*, 1998] W. E. Harris, P. R. Durell, M. J. Pierce, and J. Seckers. Constraints in the Hubble constant from observations of the brightest red giant stars in a Virgo-cluster galaxy. *Nature*, 395:45, 1998.

[Harrison, 2000] E. R. Harrison. *Cosmology: The Science of the Universe*. Cambridge University Press, Cambridge, 2 edition, 2000.

[Hartle, 2004] J. B. Hartle. Anthropic reasoning and quantum cosmology. In Allen, R. and Pope, C., editors, *The New Cosmology*. AIP, 2004. gr-qc/0406104.

[Harvey, 2006] J. Harvey. this vol, Ch. 14, 2006.

[Harwit, 1984] M. Harwit. *Cosmic Discovery*. MIT Press, 1984.

[Hawking, 1987] S. W. Hawking. Quantum cosmology. In Hawking, S. W. and Israel, W., editors, *300 Years of Gravitation*, page 631. Cambridge University Press, Cambridge, 1987.

[Hawking, 1993] S. W. Hawking. *On the Big Bang and Black Holes*. World Scientific, 1993.

[Hawking and Ellis, 1973] S. W. Hawking and G. F. R. Ellis. *The Large Scale Structure of Space-time*. Cambridge University Press, 1973.

[Hilbert, 1964] D. Hilbert. On the infinite. In Benacerraf, P. and Putnam, H., editors, *Philosophy of Mathematics*, pages 134–151. Englewood Cliff, N J.: Prentice Hall, 1964.

[Hogan, 1999] C. J. Hogan. Where the baryons are. astro-ph/9912107, 1999.

[Hogan, 2003] C. J. Hogan. Why the universe is just so. *Reviews of Modern Physics*, 72:1149–1161, 2003.

[Hollands and Wald, 2002] S. Hollands and R. M. Wald. An alternative to inflation. *Gen Rel.Grav*, 34:2043–2055, 2002.

[Hoyle, 1960] F. Hoyle. Cosmological tests of gravitational theories. In *Rendiconti Scuola Enrico Fermi XX Corso*, pages 141–. Academic Press, New York, 1960.

[Hu and Sugiyama, 1995a] W. Hu and N. Sugiyama. Anisotropies in the CMB: An analytic approach. *Astrophysical Journal*, 444:489–506, 1995.

[Hu and Sugiyama, 1995b] W. Hu and N. Sugiyama. Toward understanding CMB anisotropies and their implications. *Physical Review D*, 51:2599–2630, 1995.

[Hubble, 1936] E. Hubble. *The Realm of the Nebulae.* Yale University Press, 1936.

[Hume, 1993] D. Hume. *Dialogues and Natural History of Religion.* World Classics. Oxford University Press, 1993.

[Isham, 1997] C. J. Isham. *Lectures on Quantum Theory: Mathematical and Structural Foundations.* Imperial College Press, London, 1997.

[Jain and Dev, 2005] D. Jain and A. Dev. Age of high redshift objects - a litmus test for dark energy models. astro-ph/0509212, 2005.

[Kamionkowski and Loeb, 1997] M. Kamionkowski and A. Loeb. Getting around cosmic variance. *Physical Review D*, 56:4511, 1997.

[Kantowski, 1998] R. Kantowski. The effects of inhomogeneities on evaluating the mass parameter ω_m and the cosmological constant λ. *Astrophysical Journal*, 507:483–496, 1998.

[Kantowski *et al.*, 1995] R. Kantowski, T. Vaughan, and D. Branch. The effects of inhomogeneities on evaluating the deceleration parameter. *Astrophysical Journal*, 447:35–42, 1995.

[Khoury *et al.*, 2001] J. Khoury, B. A. Ovrut, P. J. Steinhardt, and N. Turok. The ekpyrotic universe: Colliding branes and the origin of the hot big bang. *Physical Review D*, 64:123522, 2001.

[Knobe *et al.*, 2005] J. Knobe, K. D. Olum, and A. Vilenkin. Philsophical implications of inflationary cosmology. phys/0320071, 2005.

[Kolb and Turner, 1990] E. W. Kolb and M. S. Turner. *The Early Universe.* Addison Wesley, 1990.

[Kragh, 1996] H. Kragh. *Cosmology and Controversy.* Princeton University Press, Princeton, 1996.

[Kristian and Sachs, 1966] J. Kristian and R. K. Sachs. Observations in cosmology. *Astrophysical Journal*, 143:379, 1966.

[Kuhn, 1977] T. Kuhn. Objectivity, value judgment, and theory choice. In *The Essential Tension*, pages 320–39. University of Chicago Pressm 1977.

[Lachièze *et al.*, 1995] M. M. R. Lachièze and J. P. Luminet. Cosmic topology. *Physics Reports*, 254:135, 1995.

[Lakatos, 1980] I. Lakatos. *The Methodology of Scientific Research Programmes.* Cambridge University Press, 1980.

[Landsman, 2006] N. P. Landsman. this vol, ch.5, 2006.

[Langlois and Vernizzi, 2005] D. Langlois and F. Vernizzi. Conserved non-linear quantities in cosmology. astro-ph/0509078, 2005.

[Lemaître, 1931] G. Lemaître. The beginning of the world from the point of view of quantum theory. *Nature*, 127:706, 1931.

[Leslie, 1989] J. Leslie. *Universes.* Routledge, 1989.

[Leslie, 1994] J. Leslie. Cosmology: a philosophical survey. *Philosophia*, 24:3–27, 1994.

[Leslie, 1998] J. Leslie, editor. *Modern Cosmology and Philosophy.* Prometheus Books, 1998.

[Lewis, 1986] D. K. Lewis. *On the Plurality of Worlds.* Basil Blackwell, Oxford, 1986.

[Liddle, 2004] A. R. Liddle. How many cosmological parameters? astro-ph/0401198, 2004.

[Lidsey *et al.*, 1997] J. E. Lidsey, A. R. Liddle, E. W. Kolb, E. J. Copeland, T. Barriero, and M. Abney. Reconstructing the inflation potential—an overview. *Reviews of Modern Physics*, 69,:373–410, 1997.

[Linde, 1990] A. D. Linde. *Particle Physics and Inflationary Cosmology.* Harwood Academic, 1990.

[Lineweaver and Davis, 2005] C. H. Lineweaver and T. M. Davis. Misconceptions about the big bang. *Scientific American*, 2005.

[Luminet, 2005] J.-P. Luminet. A cosmic hall of mirrors. *Physics World*, 18:23–28, 2005.

[Luminet *et al.*, 2003] J.-P. Luminet, J. R. Weeks, A. Raizuelo, R. Lehoucq, and J. P. Uzan. Dodecahedral space topology as an explanation for weak wide-angle temperature correlations in the cosmic microwave background. *Nature*, 425:593–595, 2003.

[Madsen and Ellis, 1988] M. S. Madsen and G. F. R. Ellis. Evolution of Ω in inflationary universes. *Monthly Notice of the Royal Astronomical Society*, 234:67–77, 1988.

[Malament, 2006] D. Malament. This vol ch. 3, 2006.

[Matravers *et al.*, 1995] D. R. Matravers, G. F. R. Ellis, and W. R. Stoeger. Complementary approaches to cosmology: Relating theory and observations. *Quarterly Journal of the Royal Astronomical Society*, 36:29–45, 1995.

[McCrea, 1953] W. H. McCrea. Cosmology. *Reports on Progress in Physics*, 16:321–363, 1953.

[McCrea, 1960] W. H. McCrea. The interpretation of cosmology. *Nature*, 186:1035, 1960.

[McCrea, 1970] W. H. McCrea. A philosophy for big bang cosmology. *Nature*, 228:21, 1970.

[Meyer, 1994] D. M. Meyer A distant space thermometer. *Nature*, 371:13, 1994.

[Misner, 1968] C. W. Misner. The isotropy of the universe. *Astrophysical Journal*, 151:431–458, 1968.

[Misner, 1969] C. W. Misner. Mixmaster universe. *Physical Review Letters*, 22:1071–1074, 1969.

[Morowitz, 2002] L. Morowitz. *The emergence of everything: How the world came to be complex.* Oxford, 2002.

[Mouri and Taniguchi, 2005] H. Mouri and Y. Taniguchi. Hierarchical object formation in th epeculiar velocity field. astro-ph/0508069, 2005.

[Mulryne *et al.*, 2005] D. J. Mulryne, R. Tavakol, J. E. Lidsey, and G. F. R. Ellis. An emergent universe from a loop. *Physical Review D*, 71:123512, 2005.

[Munitz, 1962] M. K. Munitz. The logic of cosmology. *British Journal for the Philosophy of Science*, 13:34–50, 1962.

[Mustapha *et al.*, 1998] N. Mustapha, C. Hellaby, and G. F. R. Ellis. Large scale inhomogeneity vs source evolution: Can we distinguish them? *Monthly Notice of the Royal Astronomical Society*, 292:817–830, 1998.

[Nilsson *et al.*, 1991] J. S. Nilsson *et al.*, editors. *The Birth and Early Evolution of our Universe: Proceedings of Nobel Symposium 19.* World Scientific, 1991.

[North, 1965] J. D. North. *The Meaure of the Universe.* Oxford University Press, 1965.

[Padmanbhan, 1993] T. Padmanbhan. *Structure formation in the universe.* Cambridge University Press, 1993.

[Partridge, 1995] B. Partridge. *3K: The Cosmic Microwave Background Radiation.* Cambridge University Press, 1995.

[Peacock, 1999] J. A. Peacock. *Cosmological Physics.* Cambridge University Press, 1999.

[Peebles, 1971] P. J. E. Peebles. *Physical Cosmology.* Princeton University Press, Princeton, 1971.

[Peebles, 1993a] P. J. E. Peebles. *The Large-Scale Structure of the Universe.* Princeton University Press, 1993.

[Peebles, 1993b] P. J. E. Peebles. *Principles of Physical Cosmology.* Princeton University Press, 1993.

[Peebles *et al.*, 1991] P. J. E. Peebles, D. N. Schramm, E. L. Turner, and R. G. Kron. The case for the relativistic hot big bang cosmology. *Nature*, 352:769–776, 1991.

[Penrose, 1989a] R. Penrose. Difficulties with inflationary cosmology. In Fergus, E., editor, *14th Texas Symposium Relativistic Astrophysics.* New York Academy of Science, New York, 1989.

[Penrose, 1989b] R. Penrose. *The Emperor's New Mind.* Oxford University Press, 1989.

[Penrose, 2004] R. Penrose. *The Road to Reality.* Jonathan Cape, 2004.

[Perez *et al.*, 2005] A. Perez, H. Sahlmann, and D. Sudarsky. On the quantum origin of the seeds of cosmic structure. gr-qc/0508100, 2005.

[Perkins, 2005] D. Perkins. *Particle Astrophysics.* Oxford, 2005.

[Perlmutter *et al.*, 1998] S. Perlmutter *et al.* Discovery of a supernova explosion at half the age of the universe. *Nature*, 391:51, 1998.

[Pettini, 1999] M. Pettini. Element abundances at high redshifts. astro-ph/9902173, 1999.

[Rees, 1995] M. J. Rees. *Perspectives in Astrophysical Cosmology.* Cambridge University Press, 1995.

[Rees, 1999] M. J. Rees. *Just six numbers: The Deep Forces that shape the universe.* Weidenfeld and Nicholson, London, 1999.

[Rees, 2003] M. J. Rees. *Our Cosmic Habitat.* Princeton University Press, Princeton, 2003.

[Rindler, 1956] W. Rindler. Visual horizons in world models. *Monthly Notice of the Royal Astronomical Society*, 116:662, 1956.

[Rindler, 2001] W. Rindler. *Relativity: Special, General, and Cosmological.* Oxford University Press, 2001.

[Robertson, 1933] H. P. Robertson. Relativistic cosmology. *Reviews of Modern Physics*, 5:62–90, 1933.

[Robertson, 1935] H. P. Robertson. Kinematics and world structure. *Astrophysical Journal*, 82:284–301, 1935.

[Rothman and Ellis, 1993] T. Rothman and G. F. R. Ellis. Lost horizons cosmology. *American Journal of Physics*, 61:883–893, 1993.

[Rovelli, 2004] C. Rovelli. *Quantum Gravity.* Cambridge University Press, 2004.

[Rovelli, 2006] C. Rovelli. This vol ch.12, 2006.

[Rugh and Zinkernagel, 2002] S. E. Rugh and H. Zinkernagel. The quantum vacuum and the cosmological constant problem. *Studies in History and Philosophy of Modern Physics*, 33:663–705, 2002.

[Saini *et al.*, 2000] T. D. Saini, S. Raychaudhury, V. Sahni, and A. A. Starobinsky. Reconstructing the cosmic equation of state from supernova distances. *Physical Review Letters*, 85:1162–1165, 2000.

[Sandage, 1961] A. Sandage. The ability of the 200-inch telescope to distinguish between selected world models. *Astrophysical Journal*, 133:355–392, 1961.

[Sandage *et al.*, 1993] A. Sandage, R. G. Kron, and M. S. Longair. *The Deep Universe: Saas-Fee Advanced Course 23.* Springer, Berlin, 1993.

[Schneider *et al.*, 1992] P. Schneider, J. Ehlers, and E. E. Falco. *Gravitational Lenses.* Springer, Berlin, 1992.

[Schramm and Turner, 1998] D. N. Schramm and M. S. Turner. Big-bang nucleosynthesis enters the precision era. *Reviews of Modern Physics*, 70:303–318, 1998.

[Schrödinger, 1956] E. Schrödinger. *Expanding Universes.* Cambridge University Press, Cambridge, 1956.

[Sciama, 1971] D. W. Sciama. Astrophysical cosmology. In Sachs, R. K., editor, *General Relativity and Cosmology*, pages 183–236. Academic Press, New York, 1971.

[Scott, 2005] D. Scott. The standard cosmological model. astro-ph/0510731, 2005.

[Seife, 2000] C. Seife. *Zero: the Biography of a Dangerous Idea.* Penguin, 2000.

[Seife, 2003] C. Seife. *Alpha and Omega: The Search for the Begining and end of the Universe.* Bantam Books, London, 2003.

[Seife, 2004] C. Seife. Physics enters the twilight zone. *Science*, 305:464–466, 2004.

[Sievers *et al.*, 2005b] J. L. Sievers *et al.* Implications of the cosmic background imager polarization data, 2005. astro-ph/0509203.

[Sigurdson and Furlanetto, 2005] K. Sigurdson and S. R. Furlanetto. Measuring the primordial deuterium abundance during the cosmic dark ages. astro-ph/0505173.

[Silk, 1997] J. Silk. *A Short History of the Universe.* Scientific American Library, 1997.

[Silk, 2001] J. Silk. *The Big Bang.* Freeman, 3 edition, 2001.

[Silk, 2005] J. Silk. *On the shores of the unknown: A short history of the universe.* Cambridge University Press, 2005.

[Smith, 1982] R. W. Smith. *The Expanding Universe: Astronomy's Great Debate.* Cambridge University Press, 1982.

[Smolin, 1992] L. Smolin. Did the universe evolve? *Classical and Quantum Gravity*, 9:173–191, 1992.

[Spergel *et al.*, 2003] D. N. Spergel *et al.* First year Wilkinson microwave anisotropy probe (WMAP) observations: Determination of cosmological parameters. *Astrophysical Journal Supplement Series*, 148:175–194, 2003. astro-ph/0302209.

[Starkman and Schwarz, 2005] G. D. Starkman and D. J. Schwarz. Is the universe out of tune? *Scientific American*, pages 48–55, 2005.

[Steinhardt, 1995] P. J. Steinhardt. Cosmology confronts the cosmic microwave background. *International Journal of Modern Physics*, 10:1091–1124, 1995. astro-ph/0407329.

[Stoeger *et al.*, 1995] W. Stoeger, R. Maartens, and G. F. R. Ellis. Proving almost-homogeneity of the universe: an almost-Ehlers, Geren and Sachs theorem. *Astrophysical Journal*, 443:1–5, 1995. gr-qc/0508100.

[Stoeger *et al.*, 2004] W. R. Stoeger, G. F. R. Ellis, and U. Kirchner. Multiverses and cosmology: Philosophical issues. astro-ph/0407329, 2004.

[Susskind, 2003] L. Susskind. The anthropic landscape of string theory. hep-th/0302219, 2003.

[Susskind, 2005] L. Susskind. *The Cosmic Landscape: String Theory and the Illusion of Intelligent Design*. Little Brown, New York, 2005.

[Susskind and Lindesay, 2004] L. Susskind and J. Lindesay. *An Introduction To Black Holes, Information And The String Theory Revolution: The Holographic Universe*. World Scientific, 2004.

[Tegmark, 1998] M. Tegmark. Is 'the theory of everything' merely the ultimate ensemble theory? *Annals of Physics*, 270:1–51, 1998.

[Tegmark, 2002] M. Tegmark. Measuring spacetime: from big bang to blackholes. *Science*, 296:1427–1433, 2002. astro-ph/0207188.

[Tegmark, 2003] M. Tegmark. Parallel universes. In J. D. Barrow, P. C. W. D. and Harper, C., editors, *Science and Ultimate Reality: Quantum Theory, Cosmology, and Complexity*, pages 459–491. Cambridge University Press, 2003. astro=ph/0302131.

[Tegmark *et al.*, 2004] M. Tegmark (the SDSS collaboration). Cosmological parameters from SDSS and WMAP. *Physical Review D*, 69:103501, 2004.

[Tegmark *et al.*, 2003] M. Tegmark, A. Vilenkin and L. Pogosan. Anthropic predictions for neuturinos masses. astro-ph/0304536, 2003.

[Tipler *et al.*, 1980] F. J. Tipler, C. J. S. Clarke, and G. F. R. Ellis. Singularities and horizons: A review article. In Held, A., editor, *General Relativityand Gravitation*. Plenum Press, 1980.

[Tolman, 1934] R. C. Tolman. *Relativity, Thermodynamics, and Cosmology*. Clarendon Press, Oxford, 1934. Dover Publications, 1987.

[Tryon, 1973] E. P. Tryon. Is the universe a quantum fluctuation? *Nature*, 246:396, 1973.

[Uggla *et al.*, 2003] C. Uggla, H. van Elst, J. Wainwright, and G. F. R. Ellis. The past attractor in inhomogeneous cosmology. *Phys Rev.*, D68:103502 (1-22), 2003.

[Uffink, 2006] J. Uffink. This vol Ch 9, 2006.

[Vilenkin, 1982] A. Vilenkin. Creation of universes from nothing. *Physics Letters B*, 117:25, 1982.

[Wainright *et al.*, 1998] J. Wainright, A. A. Coley, G. F. R. Ellis, and M. Hancock. On the isotropy of the universe: Do Bianchi VIIh universes isotropize? *Classical and Quantum Gravity*, 15:331, 1998.

[Wainwright and Ellis, 1996] J. Wainwright, and G. F. R. Ellis. *The dynamical systems approach to cosmology*. Cambridge University Press, 1996.

[Wald, 2005] R. M. Wald. The arrow of time and the initial conditions for the universe. gr-qc/0507094, 2005.

[Wald, 1983] R. M. Wald. Asymptotic behavior of homogeneous cosmological models in the presence of a positive cosmological constant. *Physical Review*, 28:2118, 1983.

[Walker, 1936] A. G. Walker. On Milne's theory of world structure. *Proceedings of the London Mathematical Society*, 42:90, 1936.

[Walker, 1944] A. G. Walker. Completely symetric spaces. *Journal of the London Mathematical Society*, 19:219–226, 1944.

[Weeks *et al.*, 2003] J. Weeks, J.-P. Luminet, A. Riazuelo, and R. Lehoucq. Well-proportioned universes suppress CMB quadrupole. astro-ph/0312312, 2003.

[Weinberg, 1989] S. W. Weinberg. The cosmological constant problem. *Reviews of Modern Physics*, 61:1–23, 1989.

[Weinberg, 1972] S. W. Weinberg. *Gravitation and Cosmology*. Wiley, New York, 1972.

[Weinberg, 2000a] S. W. Weinberg. The cosmological constant problem. astro-ph/0005265, 2000.

[Weinberg, 2000b] S. W. Weinberg. A priori distribution of the cosmological constant. *Physical Review D*, 61:103505. astro-ph/0002387, 2000.

[Wheeler, 1968] J. A. Wheeler. *Einstein's Vision*. Springer, Berlin, 1968.

[White *et al.*, 1993] M. White, L. M. Krauss, and J. Silk. Cosmic variance in CMB anisotropies: From 1° to COBE. *Astrophysical Journal*, 418:535–543, 1993.

[Will, 1979] C. M. Will. The confrontation between gravitational theory and experiment. In Hawking, S. W. and Israel, W., editors, *General Relativity: an Einstein Centenary Survey*, pages 24–89. Cambridge University Press, 1979.

[Wilson, 1989] C. Wilson. *Leibniz's Metaphysics: A Historical and Comparative Study*. Princeton University Press, 1989.

[Wright, 2006] E. L. Wright. A century of cosmology. Astro-ph/0603750.

[Yoshimura, 1988] M. Yoshimura. The universe as a laboratory for high energy physics. In Fang, L.-Z. and Zee, A., editors, *Cosmology and Particle Physics*, pages 293–340. Gordon and Breach, 1988.

[Zeh, 1992] H. D. Zeh. *The Physical Basis of the direction of Time*. Springer Verlag, Berlin, 1992.

[Zinkernagel, 2002] H. Zinkernagel. Cosmology, particles, and the unity of science. *Studies in History and Philosophy of Modern Physics*, 33:493–516, 2002.

Issues in the Philosophy of Cosmology

SUMMARY TABLE OF ISSUES AND THESES

Issue A: The uniqueness of the universe

Thesis A1: The universe itself cannot be subjected to physical experimentation

Thesis A2: The universe cannot be observationally compared with other universes

Thesis A3: The concept of 'Laws of Physics' that apply to only one object is questionable

Thesis A4: The concept of probability is problematic in the context of existence of only one object

Issue B: The large scale of the Universe in space and time

Thesis B1: Astronomical observations are confined to the past null cone, and fade with distance

Thesis B2: 'Geological' type observations can probe the distant past of our past world line

Thesis B3: Establishing a Robertson-Walker geometry relies on plausible philosophical assumptions

Thesis B4: Interpreting cosmological observations depends on astrophysical understanding

Thesis B5: A key test for cosmology is that the age of the universe must be greater than the ages of stars

Thesis B6: Horizons limit our ability to observationally determine the very large scale geometry of the universe

Thesis B7: We have made great progress towards observational completeness

Issue C: The unbound energies in the early universe

Thesis C1: The Physics Horizon limits our knowledge of physics relevant to the very early universe

Thesis C2: The unknown nature of the inflaton means inflationary universe proposals are incomplete

Issue D: Explaining the universe — the question of origins

Thesis D1: An initial singularity may or may not have occurred

Thesis D2: Testable physics cannot explain the initial state and hence specific nature of the universe

Thesis D3: The initial state of the universe may have been special or general

Issue E: The Universe as the background for existence

Thesis E1: Physical laws may depend on the nature of the universe

Thesis E2: We cannot take the nature of the laws of physics for granted

Thesis E3: Physical novelty emerges in the expanding universe

Issue F: The explicit philosophical basis

Thesis F1: Philosophical choices necessarily underly cosmological theory

Thesis F2: Criteria for choice between theories cannot be scientifically chosen or validated

Thesis F3: Conflicts will inevitably arise in applying criteria for satisfactory theories

Thesis F4: The physical reason for believing in inflation is its explanatory power re structure growth.

Thesis F5: Cosmological theory can have a wide or narrow scope of enquiry

Thesis F6: Reality is not fully reflected in either observations or theoretical models

Issue G: The Anthropic question: fine tuning for life

Thesis G1: Life is possible because both the laws of physics and initial conditions have a very special nature

Thesis G2: Metaphysical uncertainty remains about ultimate causation in cosmology

Issue H: The possible existence of multiverses

Thesis H1: The Multiverse proposal is unprovable by observation or experiment

Thesis H2: Probability-based arguments cannot demonstrate the existence of multiverses

Thesis H3: Multiverses are a philosophical rather than scientific proposal

Thesis H4: The underlying physics paradigm of cosmology could be extended to include biological insights

Issue I: The natures of existence

Thesis I1: We do not understand the dominant dynamical matter components of the universe at early or late times

Thesis I2: The often claimed physical existence of infinities is questionable

Thesis I3: A deep issue underlying the nature of cosmology is the nature of the laws of physics.

Thesis of Uncertainty: Ultimate uncertainty is one of the key aspects of cosmology

This page intentionally left blank

QUANTUM GRAVITY

Carlo Rovelli

1 INTRODUCTION

Quantum gravity (QG) is the problem of finding a theory that describes the quantum effects on gravity. These effects escape the currently accepted physical theories. Our present knowledge of the basic dynamical laws is given by quantum mechanics (QM) and quantum field theory (QFT), general relativity (GR), and the standard model of particle physics. This set of theories has obtained an empirical success nearly unique in the history of science: so far there is no evidence of observed phenomena that clearly escape or contradict this set of theories — or a minor modification of the same. But these theories become meaningless in the regimes where relativistic quantum gravitational effects are expected to become relevant. These effects are not currently observed; they are negligible at currently accessible scales and are expected to become relevant only in extreme physical regimes. For instance, they should govern the end of the evaporation of black holes, the beginning of the life of the Universe near the Big Bang, and any measurement involving an extremely short length scale ($\sim 10^{-33}$ cm, the "Planck scale") or a very high energy. "Quantum gravity" is the name given to the theory-to-be-found that should describe these regimes.

The interest of the problem, however, goes far beyond the description of some so far inaccessible physics. The physics of the early twentieth century has modified the roots of our understanding of the physical world. It has changed the very meaning of the concepts we use to grasp it. GR, which is the field theory that describes gravity when we can disregard its quantum properties, has changed our understanding of space and time. QM, which has replaced classical mechanics as our general theory of motion, has modified the notions of matter, field, and causality. At present, we haven't yet found a consistent conceptual frame in which these modifications make sense together. Thus, our understanding of the physical world is currently badly fragmented. In spite of its empirical effectiveness, fundamental physics is in a phase of deep conceptual confusion. The problem of QG is to combine the insights of GR and QM into a conceptual scheme in which they can coexist. It is the problem of finding a novel picture of the world capable of bringing the twentieth century scientific revolution to an end. For this reason, many consider QG to be the most important open problem in fundamental physics.

In particular, QG is an investigation on the nature of space and time. The structure and the nature of physical space are expected to change radically at

Handbook of the Philosophy of Science. Philosophy of Physics
Volume editors: Jeremy Butterfield and John Earman
General editors: Dov M. Gabbay, Paul Thagard and John Woods
© 2007 Elsevier B.V. All rights reserved.

the Planck scale; and the conventional way of conceptualizing of time evolution is expected to cease to be viable at this scale. The theory is therefore likely to require us to revise the way we think of space and time.

Research in QG has developed slowly for several decades of the twentieth century, because general relativity had little impact on the rest of physics and the interest of many physicists was concentrated on the development of quantum theory and particle physics. In the last decade, the explosion of empirical confirmations and concrete astrophysical, cosmological and even technological applications of general relativity on the one hand, and the satisfactory solution of most of the particle physics puzzles in the context of the particle physics "standard model" on the other, have led to a strong concentration of interest in quantum gravity, and the progressed has become rapid. Research is currently very active.

A few tentative theories of QG have been proposed. The best developed of these are string theory [Green *et al.*, 1987] and loop quantum gravity [Rovelli, 2004]. Other very active directions of investigation include noncommutative geometry [Connes, 1994], dynamical triangulations [Ambjorn *et al.*, 1997], the spinfoam formalism [Perez, 2003] (strictly connected to the loop approach) and effective theories. Currently, none of these approaches has found any empirical corroboration, and none has won general theoretical consensus.

The problem of QG raises basic methodological issues and involves conceptual and foundational questions. Some of these are similar to the foundational questions that physics addressed at the time of other major conceptual shifts — the birth of classical mechanics, field theory, relativity, or quantum mechanics. Old problems demand new answers, in the light of the twentieth century's novel insights. A characteristic example is a revival of the cartesian-newtonian-leibnizian debate on the relational nature of space.

No exhaustive discussion on our current understanding of the physical world, and in particular on the current knowledge about space and time, can disregard the issues and questions raised by this search.

1.1 Quantum spacetime

GR and QM have widely extended our understanding of the physical world. They are solidly supported empirically, and have vast scientific and technological applications. But they have destroyed the coherent picture of the world provided by prerelativistic classical physics because each of the two is formulated under assumptions contradicted by the other theory. QM is formulated using an external time variable, the t of the Schrödinger equation — or, in the case of QFT, using a fixed, nondynamical background spacetime. Both an external time variable and a fixed background spacetime are incompatible with GR.

In turn, GR is formulated in terms of Riemannian geometry: the gravitational field is assumed to be a classical deterministic dynamical field, which can be identified with Riemann's metric field. But QM requires all dynamical fields to have quantum properties. At small scales a field appears as made up of discrete quanta and is governed by probabilistic laws.

Thus, GR and QM are formulated in terms of mutually contradictory assumptions. In spite of their empirical success, they offer a rather schizophrenic and confused understanding of the physical world.

Roughly speaking, we learn from GR that spacetime is a dynamical field and we learn from QM that a dynamical field is quantized. Therefore at small scales we might expect a "quantum spacetime" formed by "quanta of space", and allowing "quantum superposition of spaces". The problem of QG is to give a mathematical and conceptual meaning to such a notion of a quantum spacetime.

Some general indications about the nature of quantum spacetime, and on the problems this notion raises, can be obtained from elementary considerations based on GR and QM. The size of quantum mechanical effects is determined by Planck's constant $\hbar$. The strength of the gravitational force is determined by Newton's constant G, and the relativistic domain is determined by the speed of light c. By combining these three fundamental constants we obtain a length, called the Planck length $l_{\rm P} = \sqrt{\hbar G/c^3} \sim 10^{-33}$ cm. Quantum-gravitational effects are likely to be negligible at distances much larger than $l_{\rm P}$, because at these scales we can neglect quantities of the order of G, $\hbar$ or $1/c$. Therefore we expect that the GR description of spacetime as a Riemannian space holds at scales larger than $l_{\rm P}$ and breaks down approaching this scale, where the full structure of quantum spacetime becomes relevant. QG is therefore the study of the structure of spacetime at the Planck scale.

Simple arguments indicate that $l_{\rm P}$ may play the role of a *minimal* length, in the same sense in which c is the maximal velocity and $\hbar$ the minimal exchanged action. For instance, the Heisenberg principle requires that the position of an object of mass m can only be determined with uncertainty x satisfying $mvx > \hbar$, where v is the uncertainty in the velocity; special relativity requires $v < c$; and according to GR there is a limit to the amount of mass we can concentrate in a region of size x, given by $x > Gm/c^2$, after which the region itself collapses into a black hole, removing itself from our observation. Combining these inequalities we obtain $x > l_P$; that is, gravity, relativity and quantum theory, taken together, appear to prevent position to be determined more precisely than the Planck scale. Various considerations of this kind have suggested that space might not be infinitely divisible. It may have a quantum granularity at the Planck scale, analogous to the granularity of the energy in a quantum oscillator. This granularity of space is fully realized in certain QG theories, such as loop quantum gravity, and there are hints of it also in string theory [Amati *et al.*, 1989]. Since this is a quantum granularity, it escapes the traditional objections to the atomic nature of space.

Time is affected even more radically by the quantization of gravity. In conventional QM time is treated as an external parameter and transition probabilities change in time. In GR there is no external time parameter. Coordinate time is a gauge variable which is not observable, and the physical variable measured by a clock is a complicated function of the gravitational field. Fundamental equations of QG might therefore not be written as evolution equations in an observable time variable. Strictly speaking, this is already true in classical GR: GR does not describe evolution of physical variables in time — it describes the relative evolution of physical variables with respect to one another. But a temporal interpretation is still available in classical GR, because spacetime appears as a solution of the dynamical equations of the gravitational field. However, a solution of the dynamical equation is like a "trajectory" of a particle and in quantum theory there are no physical trajectories: there are only transition probabilities between observable eigenvalues. Therefore in QG it may be impossible to describe the world in terms of a spacetime, in the same sense in which the motion of a quantum electron cannot be described in terms of a single trajectory. It is possible that to make sense of the world at the Planck scale, and to find a consistent conceptual framework for GR and QM, we may have to give up the notion of time altogether, and learn ways to describe the world in atemporal terms. Time might be a useful concept only within an approximate description of the physical reality.

The following section sketches the historical development of QG research and illustrates the main ideas, lines of research and current tentative theories. Various issues raised by this research are then illustrated in Section 3. Section 4 discusees in particular the changes in the notions of space and time forced by QG and Section 5 discusses the relation between the prolem of QG and other major open problems in fundamental phyiscs.

2 APPROACHES

A full account of the numerous ideas and approaches towards quantum gravity is outside the scope of this article. Only a few main research lines are illustrated here. For additional references, see the bibliographical note at the end of the article.

2.1 History and directions of research

Early ideas

The fact that the gravitational field should have quantum properties, and therefore we need a theory for describing these properties, was recognized very early. Already in 1916, one year after the birth of GR, Einstein pointed out that quantum effects must lead to modifications of GR [Einstein, 1916]. In 1927 Oskar Klein suggested that QG might ultimately modify the concepts of space and time [Klein, 1927].

In the early thirties Rosenfeld [Rosenfeld, 1930b; Rosenfeld, 1930a] wrote the first technical papers on QG, soon followed by Fierz and Pauli [Fierz, 1939; Pauli and Fierz, 1939] and later Gupta [Gupta, 1952]. The idea is to introduce a fictitious "flat space", to consider the small fluctuations of the metric around it — gravitational waves moving on flat space, described by the linearized Einstein equations — and to quantize these waves following the methods that had worked for the electromagnetic field. More precisely, the metric field $g_{\mu\nu}(x)$, which in Einstein's theory represents at the same time the spacetime metric and the gravitational field, is written as the sum of the two terms

$$(1) \quad g_{\mu\nu}(x) = \eta_{\mu\nu}(x) + h_{\mu\nu}(x).$$

$\eta_{\mu\nu}(x)$ is interpreted as the metric of a fixed background spacetime; $h_{\mu\nu}(x)$ is interpreted as the gravitational field, and quantized. A Hilbert space of states representing quantum states of gravitational waves is introduced, where $h_{\mu\nu}(x)$ is represented by a field operator $\hat{h}_{\mu\nu}(x)$. This is called the "covariant approach" to QG. The quantum of the field $h_{\mu\nu}(x)$, which is the gravitational analog of the photon, is called the "graviton", a name already in use in the early thirties.

In 1938, Heisenberg pointed out that the fact that the gravitational coupling constant is dimensional is likely to cause problems with the quantum theory of the gravitational field [Heisenberg, 1938]. In the mid thirties a young Russian physicist, Matvei Petrovich Bronstein, realized that the unique features of gravitation require a special treatment, when the full nonlinear theory is taken into account. He realized that field quantization techniques must be generalized in such a way as to be applicable in the absence of a background spacetime. Bronstein understandood that the limitation posed by GR on the mass density radically distinguishes the theory from quantum electrodynamics and would ultimate lead to the need to "reject Riemannian geometry" and perhaps also to "reject our ordinary concepts of space and time" [Bronstein, 1936].

A second line of investigation was opened in the forties by Peter Bergmann and his group [Bergmann, 1949a; Bergmann, 1949b; Bergmann, 1958; Bergmann, 1961; Bergmann and Komar, 1980]. The idea is to study and quantize the *hamiltonian* formulation of full GR, not just its linearization around flat space. This approach has the advantage that it does not assume a background spacetime on which to define the theory. The idea is that the states in the Hilbert space represent the quantum states of spacetime itself, and the full spacetime metric (maybe up to its nondynamical components) becomes a quantum operator

$$(2) \quad g_{\mu\nu}(x) \longrightarrow \hat{g}_{\mu\nu}(x).$$

This is called the "canonical approach" to QG. The same program was started independently by Dirac, who develops his constrained hamiltonian systems theory for this task [Dirac, 1950; Dirac, 1964; Dirac, 1958; Dirac, 1959].

A third approach to QG was introduced in the late fifties by Charles Misner [Misner, 1957], following a suggestion by John Wheeler. It is a quantization of general relativity à la Feynman, formally defined by the "path integral over geometries"

$$(3) \quad Z = \int Dg \ e^{-iS[g]}$$

where g is the metric field and $S[g]$ is the action of GR.

These three lines of research — covariant, canonical, and path integral — represented by equations (1), (2) and (3) respectively, still continue today. They have often influenced one another and have at times partially merged, but they have maintained a distinct flavor across more than half a century of research and are still clearly recognizable.

The basic program of the three approaches was already clearly established at the end of the fifties. The implementation of the initial programs has turned out to be a rather formidable task, but was accomplished during the sixties, in particular with the writing of the full set of Feynman rules in the covariant approach and the Wheeler-DeWitt equation in the canonical approach. Each approach, however, met serious stumbling blocks in the seventies: non-renormalizability in the covariant approach, ill-defined equations in the canonical and path integral approach. In the eighties, these stumbling blocks were overcome, in particular with the discovery of string theory in the covariant direction, and of loop quantum gravity and the (related) spinfoam formalism in the canonical and path integral directions.

The main lines of development are illustrated below.

Feynman rules and nonrenormalizability

The covariant formalism was developed during the sixties by Feynman [Feynman, 1963], DeWitt [DeWitt, 1964a; DeWitt, 1964b; DeWitt, 1965], Faddeev and Popov [Faddeev and Popov, 1967]. The technical difficulties derive from the gauge invariance of the Einstein equations, and is solved with the introduction of "ghost" particles, leading to the complete and consistent set of Feynman rules for perturbative quantum GR [DeWitt, 1967b; DeWitt, 1967c; Faddeev and Popov, 1967].

But in the early seventies the works of t'Hooft, Veltman, and then Deser and Van Nieuwenhuizen found indications that the theory does not work ['t Hooft, 1973; 't Hooft and Veltman, 1974; Deser and van Nieuwenhuizen, 1974a; Deser and van Nieuwenhuizen, 1974b], realizing Heisenberg's early fears. The reason is that the renormalization procedure, namely the technique used in QFT to remove the infinities that appear when considering the effects of arbitrarily small ("ultraviolet") field fluctuations, fails in the case of gravity. The definitive rigorous proof that the covariant quantization of general relativity fails because of nonrenormalizable ultraviolet divergences was obtained only later, in the late eighties, by Goroff and Sagnotti [Goroff and Sagnotti, 1985; Goroff and Sagnotti, 1986].

The interpretation of this failure is still controversial. There are two possibilities. One possibility is that the mistake was to assume, to start with, the existence of a background spacetime. The infinities come from short-distance fluctuations of the quantum field. These exist only if spacetime is continuous down to arbitrarily small scales. But the very fact that gravity is quantized questions the existence of such arbitrarily small scales. If, instead, spacetime has a quantized granular short-scale structure, the infinities might be just an artifact of the approximation taken in equation (1) considering $\eta_{\mu\nu}(x)$ (instead of the full $g_{\mu\nu}(x)$) as the spacetime metric. If so, the way out from the difficulty is to discard the background spacetime, and quantize the full gravitational field, as is done in the canonical or path integral approaches.

The alternative possibility is that it is GR which is not the correct theory. GR is strongly empirically supported, but only at large distances. At short distances, the world might be described by a modification of GR, with better ultraviolet behavior. There is a historical precedent: Fermi theory of weak interactions was an empirically successful but non-renormalizable theory. In the case of Fermi theory, the successful solution of the problem was to replace the theory with the Glashow-Weinberg-Salam electroweak theory, which is renormalizable and corrects Fermi theory at short scales.

Motivated by this analogy, the search for a short scale correction of GR, having better finiteness properties, has spanned several decades. After numerous attempts — some of which, like supergravity [Freedman *et al.*, 1976], and high derivative theories [Stelle, 1977] raised much hope, later disappointed — the search has lead to string theory.

Wheeler DeWitt theory

During the fifties and sixties, Bergmann's group and Dirac independently unraveled the hamiltonian structure of full GR, a rather complicated task. This structure was later clarified in the work of Arnowit Deser and Misner [Arnowitt *et al.*, 1962], using the metric $q_{ab}(x)$ of a constant-time spacelike surface, Ashtekar [Ashtekar, 1986; Ashtekar, 1987], using a connection field analogous to a Yang Mills field, and others.

In the early sixties, building on these results, Peres writes the Hamilton-Jacobi equations of GR [Peres, 1962]

$$G^2(q_{ab}q_{cd} - \frac{1}{2}q_{ac}q_{bd})\ \frac{\delta S(q)}{\delta q_{ac}}\frac{\delta S(q)}{\delta q_{bd}} + \det q\, R = 0, \tag{4}$$

were R its Ricci scalar curvature of the metric q_{ab}, and $S(q)$ is the Hamilton-Jacobi functional. In 1967, Bryce DeWitt and John Wheeler wrote the "Einstein-Schrödinger equation" [DeWitt, 1967a] following the steps taken by Schrödinger in deriving the Schrödinger equation from the Hamilton-Jacobi equation, namely interpreting the Hamilton-Jacobi equation as the eikonal approximation of a wave equation obtained replacing derivatives with derivative operators:

$$\left((\hbar G)^2(q_{ab}q_{cd} - \frac{1}{2}q_{ac}q_{bd})\frac{\delta}{\delta q_{ac}}\frac{\delta}{\delta q_{bd}} - \det q\, R\right)\ \Psi(q) = 0. \tag{5}$$

Today this is called the "Wheeler-DeWitt equation". In principle, this equation is expected to describe the full quantum dynamics of gravity. In practice, the equation remained very ill-defined for a long time: until the late eighties, when Ted Jacobson and Lee Smolin [Jacobson and Smolin, 1988] find some loop-like solutions to this equation, reformulated in Ashtekar's connection formalism, thus opening the way to loop quantum gravity.

Notice that the coordinate time variable t does not appear in the classical equation (4) nor in the quantum equation (5). This disappearance of the time variable has raised an intense debate, and will be discussed below.

Misner-Hawking sum-over-geometries

In the seventies, Steven Hawking and his group [Hawking, 1979] revived and developed the Wheeler-Misner path integral (3) in the form of a "Euclidean" integral over Riemannian (namely positive definite, as opposed to pseudo-Riemannian) metrics

$$(6) \quad Z = \int Dg \; e^{-S[g]}.$$

The hope was that the Euclidean functional integral would prove to be a better calculational tool than the Wheeler-DeWitt equation. Hartle and Hawking [Hartle and Hawking, 1983] introduced the notion of the "wave function of the universe" and the "no-boundary" boundary condition for the Hawking integral, opening up a new intuition on QG and quantum cosmology. Jim Hartle [Hartle, 1995] developed the idea of a sum-over-histories formulation of GR into a fully fledged extension of quantum mechanics to the general covariant setting. The idea was later developed and formalized by Chris Isham [Isham, 1991].

But the Euclidean integral does not provide a way of computing genuine field theoretical quantities in QG any better than the Wheeler-DeWitt equation, and the atmosphere in QG in the middle of the eighties was rather gloomy.

The idea of a sum-over-histories definition of QG was revived again in the mid nineties by the spinfoam formalism, which offer a discretized definition of the integral (3) that appears to be better defined thanks to the same short-scale spacetime discreteness implemented in loop quantum gravity.

Black hole thermodynamics

In 1974 Hawking [Hawking, 1974; Hawking, 1975] announced a theoretical derivation of black hole radiation. A (macroscopic) Schwarzschild black hole of mass M emits thermal radiation at the temperature $T = \hbar/8\pi kGM$ (k is the Boltzmann constant). The result came as a surprise, anticipated only by the observation by Bekenstein, a year earlier, that entropy is naturally associated to a black hole [Bekenstein, 1972; Bekenstein, 1973; Bekenstein, 1974] and by the Bardeen-Carter-Hawking analysis of the analogy between laws of thermodynamics and dynamical behavior of black holes [Bardeen *et al.*, 1973]. In the light of Hawking's result, the Bekenstein entropy of a Schwarzschild black hole is

$$(7) \quad S = \frac{kc^3 A}{4\hbar G}$$

where A is the area of the black-hole surface. Hawking's beautiful result is not directly connected to quantum gravity — it is a skillful application of QFT in curved spacetime, namely QFT interacting with a fixed, non quantized, gravitational field — but has a very strong impact on the field of QG. It opens a new field of research — "black hole thermodynamics" — and it opens the quantum-gravitational problems of understanding the statistical origin of the entropy (7). This is a challenge for any quantum theory of gravity.

Two years later, an influential paper by Bill Unruh [1976] convincingly argued that an observer that accelerates in the vacuum state of a conventional QFT interacts with the quantum field as if this was in a thermal bath. This shed light on black hole radiation, because an observer that remains at a fixed distance from a black hole is in constant acceleration (in order not to freely fall), and therefore black hole radiation can be interpreted simply as an Unruh effect. But at the same time this result appears to suggest that there is a deep general relation, which we do not yet understand, tying together gravity, thermodynamics and quantum theory.

In recent years, both the string and the loop approach to QG have been able to derive equation (7) from first principles [Strominger and Vafa, 1996; Rovelli, 1996a; Krasnov, 1997; Ashtekar *et al.*, 1998]. This has been considered a major success for both approaches. However, neither derivation is fully satisfactory. The string derivation does not work for conventional black holes such as a Schwarzschild black hole, but only for certain exotic black holes called extremal or nearly extremal; the loop derivation gives a finite result but the result depends on a free parameter of the theory (called the Immirzi parameter γ) that must be appropriately chosen in order to give the factor 1/4 in (7).

Noncommutative geometry

A geometrical space M admits two alternative descriptions. One is as a set of points x, the other is in terms of a commutative algebra A of functions on M. In particular, a celebrated result by Gelfand shows that a (compact Hausdorff) space M is determined by the abstract algebra A isomorphic to the algebra of the continuous functions on M. This algebraic point of view leads to a generalization of the notion of space, obtained by considering noncommutative algebras. In this sense, a noncommutative algebra defines a "noncommutative space".

Quantum theory is the discovery that the phase space of a dynamical system (the set formed by its classical states) must be replaced by a noncommutative space. In fact, the system's observables — that represent the ways we can interact with the system — form a commutative algebra of functions on the classical phase space, which becomes a noncommutative algebra in QM.

In the case of physical space, A can be identified with an algebra of coordinates, or with momentum space. If we interpret the elements of A as representing physical measurements, it is natural, in the light of quantum theory, to consider the possibility that the algebra be noncommutative. Accordingly, the hypothesis has been made that the short-scale structure of physical space might be described by a noncommutative geometry. This idea has been explored in a number of variants [Doplicher *et al.*, 1994; Doplicher *et al.*, 1995;

Doplicher, 1996].

A connection with GR has appeared in the approach developed by Alain Connes [Connes, 1994]. Connes has noticed that in the algebraic framework the notion of distance is naturally encoded in the Dirac operator D. This is the derivative operator that appears in Dirac's spinor field equation for an electron. Let $\mathcal{H}$ be the Hilbert space formed by the spinor fields on a given Riemannian (spin) manifold M, D be the (curved) Dirac operator, and A an algebra of functions on M, seen as (multiplicative) operators on $\mathcal{H}$. From the triple $(\mathcal{H}, A, D)$, called a "spectral triple", we can reconstruct the Riemannian manifold. In particular, the distance between two points x and y can be obtained as

$$(8) \quad d(x, y) = sup_{\{f\in A,||[D,f]||<1\}} \; |x(f) - y(f)|$$

a beautiful and surprising algebraic definition of distance. A non-commutative spacetime might be described by a spectral triple in which A is non-commutative. Connes suggests that this algebra may be chosen on the basis of the symmetries of the standard model, following the idea that the standard model might reveal the short-scale structure of spacetime in the same manner in which Maxwell theory revealed the structure of Minkowskian spacetime. The Connes-Chamseddine "spectral action", is simply the trace of the Dirac operator D, $S = Tr[f(D^2/(\hbar G))]$, where f is the characteristic function of the $[0, 1]$ interval. Remarkably, this action turns out to include the standard model action, including the poorly understood Higgs sector, as well as the action of GR [Chamseddine and Connes, 1996; Chamseddine and Connes, 1997]. The precise relation between the noncommutativity of noncommutative geometry and of QM has not yet been extensively investigated.

Other ideas and directions

A large number of other ideas and directions of investigation about QG have been proposed. Some of these research directions are still active. Only a few are mentioned below.

A project extensively explored is to define a quantum gravity in terms of the continuum limit of a discrete lattice theory, a technique that works in the case of quantum chromodynamics. Various attempts in this direction have failed in the past, because the lattice theory considered turned out not to have a continuum limit. One of the versions of this program, called *dynamical triangulation* is still very active, although no proof of the existence of a continuum limit exists yet.

Raphael Sorkin and his group have long explored a discrete model in which spacetime is replaced with a discrete set of points equipped with an ordering representing the causal relations [Sorkin, 1983]. Remarkably, the model has predicted a small but non-vanishing cosmological constant, of the correct order of magnitude, a prediction recently confirmed.

Roger Penrose and his group have developed twistor theory as a reformulation of metric geometry, with the hope of addressing the QG problem [Penrose, 1967]. So far, the results of twistor theory are more of mathematical than physical relevance.

Other research directions include Hartle's quantum mechanics of spacetime [Hartle, 1995], quantum Regge calculus [Williams and Tuckey, 1992; Williams, 1997], 't Hooft's deterministic approach ['t Hooft, 1996] and Finkelstein's theory [Finkelstein, 1997].

"Phenomenology" and Lorentz invariance

Until a few years ago, the research community was convinced that QG effects were certainly far outside our current observational reach. This conviction has been shaken by a number of recent suggestions that these effects might in fact be on the verge of being observable. The suggestion has even been made that certain data already observed, and which appear to be difficult to interpret with conventional physics, might be affected by QG effects. These suggestions concern for instance the cosmic propagation of high energy particles, fine details in the cosmological density spectrum, and others.

The issue appears to be related to the problem whether QG breaks Lorentz invariance. Small Lorentz noninvariant QG effects, if they exist, could be within or near observational reach. Lorentz invariant effects, on the other hand, are presumably far smaller, because Lorentz invariance forbids certain effects to happen. For instance, a small deviation from the Lorentz invariant dispersion relation governing light propagation could accumulate over cosmological travel times and yield observable frequency-dependent delays. In a Lorentz invariant context, light travel-time is a meaningless notion.

Naively one might expect that the existence of a minimal length in QG necessarily breaks Lorentz invariance. The argument is that the minimal length must be Lorentz contracted under a change of inertial frame, and therefore could not be minimal. But this argument is incorrect because it disregards quantum theory [Rovelli and Speziale, 2003]. In quantum theory a discrete quantity appears as the *eigenvalue* of an observable quantity, while a symmetry transformation transforms states, and therefore *means values*, not eigenvalues.

To illustrate this phenomenon, recall that in classical mechanics the z component L_z of the angular momentum transforms continuously under rotations. In the quantum theory, let a system be in the eigenstate $|\psi\rangle = |\hbar/2\rangle$ of L_z. Seen from a rotated reference frame, this system will appear to be in a superposition $|\psi\rangle = \alpha|\hbar/2\rangle + \beta|-\hbar/2\rangle$, where α and β vary continuously with the rotation angle. Therefore the expectation value $\overline{L_z} = |\alpha|^2\hbar/2 - |\beta|^2\hbar/2$ varies continuously in the rotation, but the eigenvalues remain the same. In physical terms: we always observe the discrete values $L_z = \pm\hbar/2$ in all reference systems — what changes continuously in a rotation is the probability of seeing one or the other. In the same fashon, in loop quantum gravity a (nonvanishing) minimal area is an eigenvalue. A surface which is in an eigenstate of the area will appear in a superposition of different area eigenstates if seeing from a boosted reference frame. The expectation value of the area of a surface can be smaller than the minimal area, but a (nonvanishing) measurement outcome cannot.

2.2 The main current tentative theories

The two currently most developed and most studied quantum theories of gravity are string theory and loop quantum gravity.

Strings

The major reason for the interest in string theory is that it is a fundamental theory of the world, including the gravitational field, which is likely to be free of ultraviolet divergences, and which encodes in a natural and strictly unified structure all the diverse ingredients we find in the world.

The starting point of the theory is the hypothesis that elementary objects are not point-like particles but rather strings, namely one-dimensional objects. The theory was initially studied as a tentative theory of the strong interactions, where it turned out to be incorrect. Quantum string theory is only consistent if spacetime has a certain dimension, called the critical dimension, which is 26 for the bosonic string and 10 for the supersymmetric string which includes fermions. The problem of reconciling the critical dimension with the fact that our world appears to be four dimensional is still open.

In 1984, Green and Schwarz introduced the idea that string theory might be a unified theory of all interactions, including gravity [Green and Schwarz, 1984]. In fact, one of the vibration modes of the proposed string has spin two, and can be identified with the graviton. Furthermore, a necessary (in general not sufficient) condition for string theory to be well-defined is that the background spacetime satisfies an equation that reduces to the Einstein equation in the large distance limit.

Consistency restricts the string models to a few alternatives. A supersymmetric model defined on a 10-dimensional flat spacetime using a large gauge group, appears to include all the ingredients of our world: the gauge group includes a subgroup which is the gauge group of the particle physics standard model, and the lowest energy vibration modes of the string include fermions, gauge bosons, and the graviton. Although no complete proof is available, the theory appears to have no ultraviolet divergences.

The idea is that six of the ten dimensions of spacetime may be invisible to us, because they are wrapped ("compactified") into a very small space (or because we are constrained to live on a four-dimensional surface). The effective physical theory in the four visible dimensions depends on the way the six extra dimensions are compactified. This can happen in a great number of different manners, giving rise to a huge number of effective four-dimensional theories. For the moment, no selection principle among this large number of possibilities has been found. Some of the resulting low energy models appear to have a strong resemblance to the standard model, but so far none seems to give precisely the physics we observe at low energy.

String theory is defined in terms of a perturbation expansion on a 10 dimensional fixed spacetime background. In the mid nineties, several nonperturbative aspects of string theory began to be investigated. Higher dimensional excitations, called "branes" (from "membrane") [Polchinski, 1995] appear to be needed in the theory for consistency, besides the strings themselves. (It has been suggested that the four-dimensional surface on which we live could be a four-dimensional brane.) The different string models appear to be related to one another (and to 11-dimensional supergravity) via simple transformations called "dualities", suggesting that all the different string models are actually different limits of a single unknown fundamental theory, tentatively-called "M-theory". The actual construction of this hypothetical fundamental theory — expected to be background independent — is still missing, and so far string theory exists only in the form of a number of (loosly) related models defined in terms of expansions over assigned background spacetimes.

In 1998, a certain conformal field theory was shown to include a sector that appears to be related to a supergravity theory on the product of Anti-deSitter spacetime and spheres. This led to the conjecture that the compactifications of string theory on an Anti-deSitter spacetimes is "dual" to a field theory on the spacetime boundary. In turn, this led to a new proposal for defining M-theory itself in term of a boundary theory: the idea is to reach background independence for M-theory using background dependent methods for the boundary theory.

The difficulties of the theory are many. No selection mechanism for the compactification is known — this is the problem of the selection of the "vacuum"; since each compactification gives different physical predictions, and there are hundreds of thousands of possible compactifications, string theory is effectively a collection of a huge number of different theories, each with different predictions and each with different physical parameters. As a result, the theory is incapable of computing the values of the standard model parameters and almost completely nonpredictive, in the sense that it can be compatible with almost any future experimental outcome. According to some critics, this lack of predictivity undermines the very nature of string theory as a scientific theory.

Even if we are willing to choose a compactification ad hoc, no compactification giving precisely the standard model in the low-energy limit is known. The theory requires supersymmetry, and the existence of observable supersymmetric particles has repeatedly been claimed as the distinctive prediction of the theory; but, in spite of several preliminary announcements, supersymmetric particles have not been found in experimental particle physics. Similarly, the possibility of detecting effects of the invisible dimensions has been considered, but experiments have given negative results. The theory requires a huge baggage of new physics (extra dimensions, an infinite number of fields with arbitrary masses and spins, supersymmetric particles . . .) but so far none of this appears to be present, or have observable consequences, in the real world.

Loops

The main reasons for interest in loop quantum gravity are: that its physical assumptions are only QM and GR, namely well-tested theories; the fact that the theory is background independent; and that it is a well developed attempt to incorporate the general relativistic notions of space and time into QFT. The theory makes no claim of being a final "Theory Of Everything". It is ultraviolet finite, without requiring high-energy modifications of GR, supersymmetry, extra dimensions, or other unobserved physics.

Loop quantum gravity was introduced in 1988. The theory is the result of the merging of two lines of research, which turn out to solve each others difficulties [Rovelli and Smolin, 1988; Rovelli and Smolin, 1990].

The first of these was the Wheeler-deWitt theory. As in the Wheeler-deWitt approach, loop quantum gravity is a straightforward quantization of GR, with its conventional matter couplings, and is based on no specific physical assumption other than GR and QM. Following the basic rules of QM, the quantum states of loop quantum gravity are obtained from a representation of an algebra of field variables of GR; their physical interpretation is obtained by diagonalizing self-adjoint operators that represent physical quantities. The difference with respect to the old Wheeler-DeWitt theory is in the choice of an algebra of loop-variables as basic variables for the quantization. Thanks to this, the ill-defined Wheeler-DeWitt theory becomes a well-defined formalism where finite physical quantities can be computed.

The second input was the idea that gauge theories are naturally described in terms of loop-like excitations. This idea can be traced back to the very beginning of field theory, an intuition of Faraday's. Faraday understood electric and magnetic phenomena in terms of lines, the "Faraday lines", that fill up space. In the presence of charges, the Faraday lines can start and end on the charges; in the absence of charges, they close, forming "loops". Maxwell translated Faraday's intuition into mathematical physics, introducing the electric and magnetic field, which are vector fields everywhere tangent to the Faraday lines, thus opening the way to modern physics, which is entirely based on the notion of field. The idea that gauge field theories are better understood in terms of loops has been defended by many scientists, including Polyakov, Mandelstam, Wilson, and others. A quantum excitation of a single Faraday line is called a "loop state".

A formulation of a QFT in terms of loop states is viable and well understood in the context of the lattice approximation; but it faces difficulties when defined over a continuum spacetime background. However, these difficulties disappear in a background independent context. The reason is that in the presence of a background, the loop states are localized on the background spacetime: there is a distinct state for each position of the loop in space. In the case of gravity, instead, there is no background spacetime. The loop states themselves are the quantum excitations of space. Therefore loop states are not immersed in space: they "weave-up" physical space themselves, in the same manner in which an ensemble of threads can weave the fabric of a T-shirt.

More precisely, the loop states of QG have self-intersection points called “nodes”. A node represents an elementary quantum excitation of space, or a single atom of space. Two nodes directly connected along a loop represent adjacent atoms of space. Nodes and links connecting nodes form a graph and carry quantum numbers. These quantum numbers determine the quantized volume of the atoms of space and the quantized area of the elementary surfaces separating adjacent nodes. A graph with these quantum numbers is called a “spin-network”, because the quantum numbers on the links turn out to be half-integers, or spins.

A spin network state does not have a position. Only combinatorial relations defining the graph are significant, not its shape or its position in space. In fact, a spin network state is not *in* space: it *is* space. Hence, in spite of its conservative basic assumptions (QM and GR), loop quantum gravity leads to a radically novel picture of space.

The possible values that the volume of a physical region or the area of a physical surface can take are determined by the spectra of the corresponding operators, following standard QM rules. These turn out to be discrete, giving the Planck-scale granular structure of space. These spectra have been computed and represent quantitative physical predictions of loop quantum gravity: a Planck-scale precision measurement of any area or volume is predicted to give as a result only the values in these spectra. For instance, the (main sequence of the) spectrum of the area is given by the expression [Rovelli and Smolin, 1995]

$$A = 8\pi\gamma\hbar G \sum_i \sqrt{j_i(j_i+1)}, \tag{9}$$

where j_i is an n-tuplet of half-integers (corresponding to the quantum numbers of the links of the spin network state crossing the surface whose area is measured). γ is the Immirzi parameter, mentioned in Section 2.1.

The dynamics is determined by a Wheeler-deWitt equation on the space of spin network states. Its ultraviolet finiteness is a consequence of the granular structure of space. Different finite and well-defined versions of this equation have been constructed. At present it is not yet clear which of these, if any, is the physically correct one.

Applications of the theory include a derivation of the Bekenstein black hole entropy mentioned in Section 2.1, applications to the description of the classical singularities, such as the ones at the center of bloack holes, and applications to cosmology. The theory appears to be capable of controlling the black hole singularities and the initial Big Bang singularity. Indirect empirical evidence supporting predictions of the theory is actively searched in the astrophysical and cosmological domains.

The main difficulties of loop quantum gravity lie in recovering low energy phenomenology. Quantum states corresponding to the Minkowski vacuum and its excitation have not yet been constructed, and particle scattering amplitudes have not been computed. This deficiency weakens the strength of the finiteness claim, and bears on one of the key requirements on a quantum theory of gravity: full recovery of low energy physics. The dynamics is still poorly understood: the Wheeler-deWitt equation exists in more than one version. The lack of unitary evolution in time and the overall radical conceptual novelty of the results of the theory, where background spacetime is discarted altogether, are questioned by some.

The loop-string debate

A theory begins to be credible only when its original predictions are reasonably unique and are confirmed by new experiments. Neither loop quantum gravity nor string theory — nor any other current tentative theory of QG — are yet credible in this sense. Furthermore, in spite of much effort, both theories are still badly incomplete and far from being clearly understood. The problem of QG must therefore be considered still fully open.

Nevertheless, in both directions the research has progressed considerably in recent years: many problems that appeared too hard ten years ago have now been solved, and incomplete but *possible* solutions of the QG puzzle are now at hand.

However, the two theories differ profoundly in their hypotheses, achievements, specific results, and in the conceptual frame they propose. The issues they raise concern the foundations of the physical picture of the world, and the debate between the two approaches involves conceptual, methodological and philosophical issues.

The lesson of string theory appears to be that in order to remove the difficulties of the perturbative quantization of GR we have to couple the gravitational field to matter. Finiteness is achieved by replacing the pointlike Feynman vertices of conventional QFT with non-point-like interactions between strings, which are extended objects. The theory preserves the basic conceptual structure of QFT (background spacetime, unitarity, predictions in terms of an asymptotic S-matrix. . .) at the prices of renouncing a full implementation of the general covariance that characterizes GR, of huge extra baggage (extra dimensions, supersymmetry, infinite fields. . .) and of a dramatic decrease in predictiveness.

Loop quantum gravity, on the other hand, is rooted in the general covariance that characterizes GR. Ultraviolet finiteness is a consequence of the granular structure of space, which, in turn, is a standard quantum mechanical effect appearing when we regard GR as a theory of spacetime itself, and not as a theory of small perturbations around a background spacetime. The interest of the loop theory, therefore, is that it is a determined effort towards a genuine merger of QFT with the world view that we have discovered with GR. Furthermore, it leads to well-defined physical predictions which are in principle falsifiable. However, even disregarding the incompleteness of the theory, the conceptual price for this result is heavy: the theory gives up unitarity, time evolution, Poincaré invariance at the fundamental level, and the very notion that physical objects are localized in space and evolve in time.

Whether these radical conceptual steps are viable, and, if viable, whether they are justified, is a hotly debated issue.

3 METHODOLOGICAL ISSUES

3.1 Justification of the quantum gravity search

Absence of empirical data

The first obvious question about the search towards QG is whether the search is legitimate at all, given the total absence of empirical data directly about the regimes QG is concerned with. We have no direct empirical guidance in searching for QG — as, say, atomic spectra guided the discovery of quantum theory.

Some critics have argued that the QG search is futile, because anything might happen in QG regimes, at scales far removed from our experience. Maybe the search is impossible because the space of possible theories is too large.

At present, this worry is probably unjustified. If this were the problem, we would have plenty of complete, predictive and coherent theories of QG, and the problem would be the choice among them. Instead, the situation is the opposite: we haven't any. The fact is that we do have plenty of information about QG, because we have QM and we have GR. Consistency with QM and GR, plus internal consistency, form an extremely strict set of constraints. The problem currently debated is to find at least one complete and consistent theory of QG. If more will be found, we will have of course to resort to experiments to select the physically correct one.

Should gravity be quantized?

The possibility that quantum gravitational effects do not exist and gravity is intrinsically classical (non quantum) has been often suggested. The justification for this suggestion is that gravity can be seen as an interaction profoundly different from the others, since it admits a description in terms of spacetime geometry. This suggestion has also generated a research program, aiming at testing the consistency of a theory in which classical gravity interacts with QFT.

In its simpler form, this suggestion has today been largely abandoned. The reason is that, as was noticed in the early days of QM, an interaction between a classical and a quantum variable is always inconsistent. If Heisenberg uncertainty relations are violated for one dynamical variable, they are violated for all other variables as well. The idea of circumventing gravity quantization, however, has reappeared under various forms.

One suggestion is that the gravitational field may not represent true microscopic degrees of freedom, but only a collective, or "hydrodynamical", large scale description of these. This hypothesis is supported by phenomena such as the relations between gravity and thermodynamics revealed by the Unruh effect. Ted Jacobson [Jacobson, 1995] has even been able to derive the Einstein equations from (7) and standard thermodynamical relations, providing evidence that could be interpreted as supporting this idea. However, even if the gravitational field is just a collective variable, this does not mean that it will not display quantum effects. QM does not govern just elementary degrees of freedom; it governs all degrees of freedom, including collective ones. Thus, this possibility, even if realized, would not refute the need of a quantum theory of gravity.

Another suggestion is that gravity may be an emergent phenomenon induced by the other quantum fields. This idea is suggested by the fact that the renormalization process for a QFT on a curved spacetime generates terms in the action which are proportional to polynomials in the Riemann curvature, and the lowest order term is precisely the action of GR. The difficulty about this suggestion is that it is ambiguous as regards the dynamical status of the metric field. The variational principle states that dynamics is determined by the variation of the action with respect to the dynamical variables only, not with respect to anything appearing in it. If the metric field is assumed to be a dynamical variable, then it is a dynamical field like any other, and the fact that the dependence of the action with respect to it is modified by the renormalization of its interaction with other fields may change the details of its dynamics, but not the fact that it is a quantum field. If, on the other hand, the metric is not a dynamical field, then the action must not be varied with respect to it (as it is not varied with respect to it in the special relativistic context), and therefore the Einstein equations are not generated by the new terms in the action. In the first case the gravitational field needs to be quantized; while the second case is in contradiction with the empirical fact that the classical Einstein equations are satisfied

3.2 Research attitudes

Different attitudes can be distinguished in the physics community with respect to the methodology used for searching for a QG theory.

(a) The "pessimistic" attitude, already mentioned above, is that of those who worry that too many possibilities are open, anything might happen between here and the Planck scale, and the search for a quantum theory of gravity is therefore futile.

As mentioned, this worry is unfounded, because we do not have too many complete QG theories: we haven't any.

(b) The view is often expressed that some totally new, radical and wild hypothesis is needed for QG. This "wild" attitude is based on the observation that great scientists had the courage to break with old and respected assumptions and to explore some novel "strange" hypotheses. From this observation, the "wild" scientist concludes that any strange hypothesis deserves to be investigated, even if it violates well established facts.

On historical grounds, this expectation is probably ill-founded. Wild ideas pulled out of the air have rarely made science advance. The radical hypotheses that physics has successfully adopted have always been reluctantly adopted because they were forced by new empirical data — Kepler's ellipses, Bohr's quantization, Planck's energy quanta — or by stringent theoretical deductions — Maxwell displacement current, Einstein's relativity. Generally, arbitrary novel hypotheses have led nowhere. This consideration leads to the next attitude in (c).

(c) Part of the research in QG is motivated by the hope that the knowledge of the world coded into GR and QM can be a good guide for finding a theory capable of describing physical regimes that we have not yet explored.

A motivation for this hope is that today we are precisely in one of the typical situations in which theoretical physics has worked at its best in the past. Many of the most striking advances in theoretical physics have derived from the effort to find a common theoretical framework for two basic and apparently conflicting discoveries. For instance, the aim of combining special relativity and non-relativistic quantum theory led to the theoretical discovery of antiparticles; combining special relativity with Newtonian gravity led to general relativity; combining the Keplerian orbits with Galilean physics led to Newton's mechanics; combining Maxwell theory with Galilean relativity led to special relativity, and so on. In all these cases, major advances have been obtained by "taking seriously" apparently conflicting theories, and exploring the implications of holding true the essential tenets of both theories. Today we are in one of these characteristic situations. We have learned two new very general "facts" about nature, expressed by QM and GR: we have "just" to figure out what they imply, taken together.

(d) A different point of view on the problem is held by those who accept that QM has been a conceptual revolution, but do not view GR in the same way. According to this point of view, the discovery of GR was "just" the writing of one more classical field theory. This field theory is likely to be only an approximation to a theory we do not yet know, and its teachings should not be overestimated. According to this opinion, GR should not be taken too seriously as a guidance for theoretical developments.

A possible objection to this point of view is that it derives from the confusion between (i) the specific form of the GR action and the GR field equations and (ii) the modification of the notions of space and time engendered by GR. The GR action could be a low energy approximation of something else. But the modification of the notions of space and time has to do with the diffeomorphism invariance and the background independence of the theory, not with its specific form. The challenge of QG is to incorporate this novelty into QFT, not the specific form of the GR action.

(e) A common attitude is the "pragmatic" attitude of the physicist who prefers to disregard or postpone these foundational issues and, instead, develop and adjust current theories. This style of research was effective during the sixties in the search for the particle physics standard model, where a long process of adjustment of existing QFT's led to a very effective theory.

It is questionable whether this attitude could be effective in a situation of foundational confusion like the present one. During the sixties empirical data were flowing in daily, to keep research on track. Today no new data are available. The "pragmatic" attitude may mislead the research: in the extreme case, the "pragmatic" physicist focuses only on the development of the theory at hand, without caring if the world predicted by the theory resembles less and less the world we see. Sometimes he is even excited that the theory looks so different from the world, thinking that this is evidence of how far ahead he has advanced in knowledge. But it is more likely that the difference between the theory and the world is only evidence of how much he is lost. Unfortunately similar excesses plague theoretical research today.

The cumulative aspect of scientific knowledge and influence of the philosophy of science

The "pessimistic", "wild" and "pragmatic" attitudes illustrated above may have been influenced by a philosophy of science that under-emphasizes the cumulative aspect of scientific knowledge, and emphasizes, instead the "incommensurability" between an old theory and a new theory that historically supersedes it. More or less informed awarness of this long standing debate in philosophy of science has indeed affected the reserach attitude of many theoreticians.

On the other hand, attitude (c) described above is based on the expectation that the central physical tenets of QM and GR represent our best guide for accessing the unexplored territories of the quantum-gravitational regime. In a difficult research situation where cataclysmatic evolution is expected anyway in the consequences of the theory (for instance, the change of the nature of space), conservative assumptions based on the confidence on the cumulative aspect of knowledge can play an important role.

This faith in a cumulative aspect of scientific knowledge is based on the idea that there are discoveries that are "forever". For instance, the Earth is not the center of the universe, simultaneity is relative, absolute velocity is meaningless, and we do not get rain by dancing.

The fact that major aspects of a theory can have value outside the domain for which the theory was discovered may be at the root of much of the historical effectiveness of theoretical physics, and in particular of spectacular predictions such as Maxwell's radio waves, Dirac's antimatter or GR's black holes. This can perhaps be understood as just scientific induction: as a consequence of the fact that Nature has regularities. This is not the place to enter this discussion; but it is relevant to remark that the existence of these regularities is held by several researchers in QG as a source of confidence — althought, of course, not certainty — that the basic facts about the world found with QM and GR will be confirmed, not violated, in the quantum gravitational regimes that we have not yet empirically probed.

4 THE NATURE OF SPACE AND TIME

GR has modified the way we understand space and time. Combining GR with QM requires a further modification of these notions. It is important, however, to clearly distinguish the modifications of the notions of space and time required by QG from the ones already implied by GR alone. These are briefly summarized in Section 4.1 below. Section 4.2 and 4.3 then discuss the notions of space and time in QG.

4.1 The physical meaning of GR

GR is the discovery that spacetime and the gravitational field are the same entity. What we call "spacetime" is itself a physical object, in many respects similar to the electromagnetic field. We can say that GR is the discovery that there is no spacetime at all. What Newton called "space", and Minkowski called "spacetime", is nothing but a dynamical object — the gravitational field — in a regime in which we neglect its dynamics.

In newtonian and special relativistic physics, if we take away the dynamical entities — particles and fields — what remains is space and time. In general relativistic physics, if we take away the dynamical entities, nothing remains. The space and time of Newton and Minkowski are reinterpreted as a configuration of one of the fields, the gravitational field. This implies that physical entities — particles and fields — are not all immersed in space, and moving in time. They do not live on spacetime. They live, so to speak, on one another.

In classical GR it is customary to maintain the expressions "space" and "time" to indicate aspects of the gravitational field. But in the quantum theory, where the field can have quantized "granular" properties and its dynamics is quantized and therefore only probabilistic, most of the "spatial" and "temporal" features of the gravitational field are probably lost.

This absence of the familiar spacetime "stage" is called the *background independence* of the classical theory. Technically, background independence is realized by the gauge invariance of the GR action under (active) diffeomorphism. A diffeomorphism is a transformation that smoothly drags all dynamical fields and particles on the four-dimensional coordinate manifold. In turn, gauge invariance under diffeomorphism (or *diffeomorphism invariance*) is the consequence of the combination of two properties of the action: its invariance under arbitrary changes of coordinates (or *general covariance*) and the fact that there is no non-dynamical "background" field. Thus: *background independence* = *diffeomorphism invariance* = (*general covariance*+absence of non-dynamical background fields). These notions are illustrated in more detail below.

Diffeomorphism invariance

Pre-general-relativistic field theories are formulated in terms of a spacetime manifold M, and a set of fields $\varphi_1, \ldots, \varphi_n$ on M. The manifold M is a (pseudo-)metric space whose points $P \in M$ represent the physical points of spacetime. Spacetime points are labelled by coordinates $x = (x^1, x^2, x^3, x^0)$ that represent the reading of measuring devices: clocks and distance-measuring devices ("rods"). More precisely, M is equipped with a (pseudo-)distance function $d(x, y)$ interpreted as the 4-interval between the two points x and y: a negative $d^2(x,y)$ gives the time measured by a clock in inertial motion between x and y; a positive $d^2(x,y)$ gives the proper length of a rod with the ends on x and y, in a state of inertial motion with respect to which x and y are simultaneous according to Einstein's definition of simultaneity; a null $d(x,y)$ indicates that light travels in vacuum from x to y. Notice that pre-general-relativistic physics deals with (relations between) two distinct types of measurements: (i) spacetime measurements measuring spacetime observables, performed by means of clocks and distance-measuring devices, and (ii) field measurements, measuring field observables, namely the values (or functions) of the fields $\varphi_1, \ldots, \varphi_n$.

This same interpretation framework is used in special-relativistic QFT. The only difference is that field observables can be quantized. The number of excited quanta has a particle interpretation. In a typical high energy scattering experiment, for instance, the field observable (ii) is the number of particles revealed by a particle detector (which is a field measuring device); while the spacetime observable (i) is the momentum of the particle, determined by measuring the spacetime position of the detector.

The theory does not predict the value of field observables φ alone, or spacetime observables x alone, but only combinations of the two, such as the value $\varphi(x)$ of a field φ at a certain spacetime location x. Spacetime and field observables are both quantities that have a direct operational interpretation; they can be called "partial observables". On the other hand, the quantities that can be predicted by the theory, such as $\varphi(x)$, for a given position x, can be called "complete observables".

The interpretation of a *general* relativistic field theory is different. In such a theory there is a field g representing the gravitational field and possibly other fields, representing other dynamical variables. These fields are defined on a differentiable manifold M, coordinatized by coordinates x. The formal structure of a general relativistic field theory is therefore similar to the structure of a pre-general-relativistic field theory. But two major differences force a different interpretation. First, the manifold M on which the fields are defined is not a metric manifold. The gravitational field g equips M with a metric structure $d_g(x,y)$.[1] Therefore clocks and distance-measuring-devices measure properties of the gravitational field g. It follows that the distinction between spacetime observables of the kind (i) and field observables of the kind (ii) is blurred. This blurring of the distinction between the two kind of partial observables is a crucial conceptual novelty of GR.

Second, the field equations are invariant under a transformation of the fields called active diffeomorphisms. An active diffeomorphism $g \to \tilde{g}$ is determined by (but should not be confused with) a smooth invertible function $f : M \to M$. Under a diffeomorphism transformation, the field g and all other fields are "dragged along" M by f. For instance, the transformed field $\tilde{g}$ defines a new distance

[1] The length of a curve $\gamma^\mu(s)$ in M is $d_g(\gamma) = \int ds \sqrt{g_{\mu\nu}(\gamma(s)) \frac{d\gamma^\mu(s)}{ds} \frac{d\gamma^\nu(s)}{ds}}$ and $d_g(x,y)$ is a local extremum of $d_g(\gamma)$ over the curves that join x and y.

function $d_{\tilde{g}}(x,y)$ which is related to the one defined by g by

(10) $d_{\tilde{g}}(f(x), f(y)) = d_g(x,y)$.

In words: the distance between the two points $f(x)$ and $f(y)$, defined by the field $\tilde{g}$ is the same as the distance between the two points x and y defined by g.

The importance of this invariance of the field equations is due to the following. An active diffeomorphism may modify a solution of the GR equations in the future of a certain time surface t_0, without modifying it at all in the past of t_0. Therefore two distinct solutions of the equations of motion can be equal in the past and differ in the future. This fact gives us a choice: either (i) we interpret the theory as an indeterministic theory, where the future is not determined by the past. Or (ii) we interpret active diffeomorphisms as a "gauge invariance": that is, we postulate that the complete observables of the theory are only given by quantities that are invariant under this transformation. The alternative (i) is not viable, because experience shows that classical gravitational physics is completely deterministic. We are therefore forced to alternative (ii), which has heavy interpretative consequences.

To understand these consequences, let P be a point of M. Let $\varphi(P)$ be any property of the fields at P. For instance, $\varphi(P)$ may represent a value of the electromagnetic field at P or the spacetime scalar curvature at P, or something similar. None of these properties is invariant under active diffeomorphism. Therefore, it follows from the argument above that none of these properties can be predicted by the theory. Therefore the theory does not determine the physics at spacetime points $P \in M$.

At first this conclusion might sound bewildering: if physics does not predict what happens at spacetime points, what can it predict? In fact, historically, Einstein himself got at first confused and frustrated by this observation, to the point of stepping back from the diffeomorphism invariance he previously expected GR to have [Norton, 1984]. Einstein's version of the argument given above is called the "hole argument" (because Einstein considered a diffeomorphism affecting only a finite region of spacetime, empty of matter, or a "hole"), and was presented in [Einstein and Grossmann, 1914]. On this argument, and the discussion it has raised, see for instance [Earman, 1987; Earman, 1989; Belot, 1998; Earman, 2001; Pauri and Vallisneri, 2002] and references therein. Later however, Einstein changed his mind and accepted both diffeomorphism invariance and the conclusion (ii), realizing that this conclusion fully implemented his intuition on the very central physical meaning of the general relativistic conceptual revolution.

The way out of the puzzle is to understand that in the general-relativistic context the points of the manifold do not represent physical entities with an existence independent from the fields. Asking what are the properties of the fields at P is meaningless. Spacetime locations can only be determined by the fields themselves, or, by any other dynamical object we are considering. For instance, if the theory we consider includes two particles and the trajectories of the two particles happen to meet once, then the meeting of the particles determines a spacetime point. The theory is able to predict the value of the fields and any other physical properties at the spacetime point determined by the meeting of the particles. However, this point cannot be naively identified as a point of M, because the same physical situation can be represented by the set of dynamical variables obtained by an active diffeomorphism, where the particles meet now at a different point, say Q of M. The value of the fields at the point where the particle meet is invariant under such a transformation, because the particles' trajectories and the fields are dragged along M together. Einstein called this way of determining location in terms of the dynamical objects (fields and particles) of the theory itself, "spacetime coincidences".

Thus, a general relativistic theory does not deal with values of dynamical quantities at given spacetime points: it deals with values of dynamical quantities at "where"'s and "when"'s determined by other dynamical quantities.

Strictly speaking also in the pre-general-relativistic context physics deals with values of dynamical quantities at "where"'s and "when"'s determined by other physical quantities, because the times and distances used to determine location are physical quantities. But in the pre-general-relativistic context we can make a strict separation between: (i) "spacetime", viewed as a background entity, and measured by clocks and rods that one considered non-dynamical, and (ii) dynamical variables. In the general-relativistic context, on the other hand, this separation is lost, time and distance measurements are reinterpreted as measurements of the gravitational field, on the same footing as other field measurements, and there is no distinction between non-dynamical background and dynamical physical variables.

Physical meaning of the coordinates

A consequence of the above is that in the general-relativistic context the physical interpretation of the coordinates x is different from their interpretation in the non-general-relativistic context. In the general-relativistic context the coordinates x have no interpretation at all: observable quantities in GR correspond to quantities of the theory that are independent of the coordinates x. Recall that the non-general-relativistic coordinates represent the reading of clocks and rods: in the general-relativistic context, the reading of clocks and rods is represented by the non-local function $d_g(x, y)$ of the gravitational field. The fact that the non-general-relativistic coordinates x and the general-relativistic coordinates x are denoted in the same manner is only an unfortunate historical accident.

To illustrate in which sense observable quantities are independent of the coordinates x, consider a typical general-relativistic measurement. A standard application of GR is in precision measurements and precision modeling of solar system dynamics. In this context, partial observable quantities are the "instantaneous" distances d_p between the Earth and the different planets, defined as the proper time elapsed on Earth (measured by a clock at rest on Earth) while a radar signal goes from Earth to a planet p and back. Fixing an arbitrary initial event on Earth, one additional partial observable can be obtained as the proper time

τ from this event along the Earth trajectory. Complete observables are then the values $d_p(\tau)$ of the planet distances, at different local proper times τ. A general relativistic model of the solar system, with appropriately chosen initial data, can predict $d_p(\tau)$ for all p's and all τ's, and these predictions can be compared with experience. In building up this model, we choose an arbitrary coordinatization x of the solar system region, and express the gravitational field, the electromagnetic field, and the planets' positions in this coordinate system. The predicted quantities $d_p(\tau)$ are complicated non-local functions of the fields and planets' positions, which are independent of the coordinates x chosen. To be sure, the observable τ is introduced here only for convenience. We can equivalently express the predictions of the theory simply as a set of relations $f(d_p) = 0$ that must hold between the partial observables d_p.

General covariance and Kretschmann objection

The invariance of GR under active diffeomorphisms follows from two properties of the GR field equations. First, they are generally covariant. That is, they maintain the same form under any smooth change of coordinates $x \to x'(x)$ on M. This means that there is no coordinate systems on M which is preferred a priori. Second, there are no fixed non-dynamical fields in the field equations.

The first property, namely general covariance, is the property that Einstein most insisted upon, and that guided him in finding GR. The requirement of general covariance still plays a major role in selecting physical theories compatible with what we have understood about the world as a result of the general relativistic revolution.

However, general covariance *alone* is not excessively significative. Indeed, any field equation can be written in an arbitrary coordinate system. This fact was pointed out by Kretschmann shortly after Einstein wrote GR [Kretschmann, 1917], and has raised much discussion. As an example, consider the field equation

$$(11)\quad (\partial_T^2 - \partial_X^2 - \partial_Y^2 - \partial_Z^2)\ \varphi(X,Y,Z,T) = 0.$$

If we introduce arbitrary coordinates x (with components x^μ) as functions $x^\mu = x^\mu(X,Y,Z,T)$ (with inverse $X(x), Y(x), Z(x), T(x)$), the wave equation (11) becomes the generally covariant equation

$$(12)\quad \Box_g \varphi(x) \equiv \partial_\mu\ \sqrt{\det -g(x)}\ g^{\mu\nu}(x)\ \partial_\nu \varphi(x) = 0,$$

In this equation, the unknown is $\varphi(x)$, while $g^{\mu\nu}(x)$ and $\det g(x)$ are the inverse and the determinant of the *fixed* field

$$(13)\quad g_{\mu\nu}(x) = \frac{\partial X(x)}{\partial x^\mu}\frac{\partial X(x)}{\partial x^\nu} + \frac{\partial Y(x)}{\partial x^\mu}\frac{\partial Y(x)}{\partial x^\nu} + \frac{\partial Z(x)}{\partial x^\mu}\frac{\partial Z(x)}{\partial x^\nu} - \frac{\partial T(x)}{\partial x^\mu}\frac{\partial T(x)}{\partial x^\nu}.$$

The field theory for the scalar field φ defined by equation (12) is *not* diffeomorphism invariant, because in distinct coordinate systems the field equations for the unknown φ are different, in the sense that they are determined by different functions $g_{\mu\nu}(x)$.

Equation (12), on the other hand, can become one of the equations of a diffeomorphism invariant theory in which $g_{\mu\nu}(x)$ is also one of the unknown, namely a dynamical field. Therefore whether or not a theory is diffeomorphism invariant is not determined just by the aspect of an equation, but by the full specification of the dynamical quantities and their equations of motion.

Eistein's insistence on general covariance alone, however, should probably not be interpreted as a lack of clarity on his part, but only as his effort to emphasize the importance of a step that was *necessary*, not sufficient, to write a succeful relativistic field theory.

In addition, like all formal properties of physical theories, even full diffeomorphism invariance, should probably not be interpreted as a rigid selection principle, capable of selecting physical theories *just by itself.* With sufficient acrobatics, any theory can perhaps be re-expressed in a diffeomorphism invariant language. The same is true for any other formal invariance property. For instance, any theory can be rewritten as a rotational invariant theory, or with any other desired invariance property, by simply adding variables.

As an example, equation (11) can be viewed as physically equivalent to the diffeomorphism invariant system

$$(14) \quad \Box_g \varphi(x) = 0, \qquad Riem[g] = 0,$$

where $Riem[g]$ is Riemann's curvature and the unknowns are now the two fields φ and g. But there are prices to pay. First, this theory has a "fake" dynamical field, since g is constrained to a single solution up to gauges, by the second equation of the system. Having no physical degrees of freedom, g is physically a fixed background field, in spite of the trick of declaring it a variable and then constraining the variable to a single solution. Second, we can insist on a lagrangian formulation of the theory (14) [Sorkin, 2002], but to do this we must introduce an additional field, and it can then be argued that the resulting theory, having an additional field is different from (12) [Earman, 1989].

Diffeomorphism invariance is the key property of the mathematical language used to express the key conceptual shift introduced with GR: the world is not formed by a fixed non-dynamical spacetime structure, which defines localization and on which the dynamical fields live. Rather, it is formed solely by dynamical fields in interactions with one another. Localization is only defined, relationally, with respect to the fields themselves.

Relationalism and substantivalism

A non-dynamical background space was used by Newton. The first part of the *Principia*, Newton argues very explicitly that we must assume the existence of space as an entity. This part can be read as a polemic against the long dominant, and in particular Descartes's, relational understanding of space.

The two traditional views about space, absolute ("space is an entity") and relational ("space is a relation between entities"), suitably modified to take into account scientific progress, continue in contemporary philosophy of science under the names of *substantivalism* and *relationalism*.

We can say that Einstein has "unmasked" the entity introduced by Newton (which much disturbed Leibniz): Newton's space is nothing else than a field like the others, thought Newton considered it in a regime in which its dynamics could be neglected. Localization in space and in time, introduced by Newton against Descartes's relational localization, is revealed by Einstein to be, after all, still a relational location — in the sense of Descartes —, with respect to a specially chosen entity: the gravitational field. In a sense, we can therefore say that GR realizes a full return to a relational definition of space and time, after the Newtonian substantivalist parenthesis.

In other words, in prerelativistic physics, spacetime is a sort of structured container which is the home of the world. In general-relativistic physics, on the other hand, there is nothing of the sort. There are only interacting fields (including the gravitational field) and particles: the only notion of localization which is present in the theory is relative: dynamical objects (fields and particles) are localized only with respect to one another. This is the notion of relational space defended by Aristotle and Descartes, against which Newton wrote the initial part of the *Principia*. Newton had two points in his favor: the physical reality of inertial effects such as the concavity of the water in the bucket of his famous bucket experiment, and the immense empirical success of his theory based on absolute space. Einstein has provided an alternative interpretation for the cause of the concavity — the interaction with an entity: the local gravitational field — and a theory based on relational space that is empirically far more effective than Newton theory. Einstein has therefore reopened the possibility of a relational understanding of space and time, which was closed by Newton's bucket.

At the basis of Cartesian relationalism was the notion of "contiguity". Two objects are contiguous if they are adjacent to one another. Space is the order of things with respect to such contiguity relation. At the basis of the spacetime structure of GR there is a very similar notion: Einstein's "spacetime coincidence" is strictly analogous to Descartes' "contiguity".

The key to this novel relational understanding of space and time is Faraday's revolutionary idea that a field is a physical entity. Recall that Faraday visualized a field as a family of real lines filling up everything. Einstein's entire theoretical work fully implements this realistic interpretation of the fields. In Einstein popular-science writing, the gravitational field is a huge "jelly fish", a better metaphor than the lines of Faraday. Entities are not just particles, but also fields, the gravitational field is one field among the others. These entities are localized only with respect to one another.

A substantivalist position can nevertheless still be –and in fact is still– defended. Einstein's discovery that Newtonian spacetime and the gravitational field, are the same entity, can also be expressed by saying that there is no gravitational field: it is spacetime that has dynamical properties. This choice is not uncommon in the literature. The difference with the laguage used here is only a matter of choice of words. The substantivalist can therefore claim use the that, according to GR, "spacetime is an entity": indeed, it is the gravitational field, which is an entity. Since it is possible to define localization with respect to the gravitational field, the substantivalist can also say that "spacetime is an entity that defines localization".

However, this is an extremely weakened substantivalist position. To what extent is general-relativistic spacetime different from any other arbitrary entity with respect to which we can define a relational localization? We can call "spacetime" anything used to define localization. Newton's acute formulation of substantivalism contains a precise characterization of "space" [Newton, 1962]:

> "...So it is necessary that the definition of places, and hence of local motion, be referred to some motionless thing such as extension alone or "space", *in so far as space is seen to be truly distinct from moving bodies.*"

(My italic.) The characterizing feature of space, according to this substantivalist manifesto, is to be "truly distinct from moving" bodies. In modern terms and after the Faraday and Maxwell conceptual revolution, I believe this can only be translated as being "truly distinct from dynamical entities such as particles or fields". This is *not* the case for the spacetime of GR. The modern substantivalist can give up Newton's strong substantivalism ("spacetime is a non-dynamical entity") for the much weaker thesis "we call spacetime the gravitational field, which is a dynamical entity". But then what is the difference between this position and the relationalist one, if not just a choice of words?

To be sure, general relativistic relationalism doesn't fit comfortably with traditional relationalism either. E.g. observables of GR, conceived as coincidence quantities are non-substantival in that they don't require spacetime points to support them. But neither are they relational in the traditional sense of involving relations between "material" bodies or events in their histories.

The traditional substantivalist-relational alternative was formulated before the Faraday-Maxwell conceptual revolution, without taking the existence of the *fields* into account. After Farady and Maxwell, we understand the world also in terms of a new set or dynamical entities, the fields. Once we accept the existence of the fields, and Einstein's discovery that Newton's space is one of the fields, the distinction between substantivalism and relationalism is largely reduced to mere semantics.

When two opposite positions in a long-standing debate have come so close that their distinction is reduced to semantics, one can perhaps say that the issue is resolved. In this sense, it may be argued that GR has solved the long-standing issue of the relational versus substantivalist interpretation of space.

4.2 Background independence

Is QM compatible with the general relativistic notions of space and time sketched above? It is, but a sufficiently general formulation of QM must be used. For instance, the Schrödinger picture is only viable for theories where there is a global observable time variable t; this conflicts with GR, where no such variable exists. Therefore the Schrödinger picture makes little sense in a background independent context. But formulations of QM have been proposed that are more general than the Schrödinger picture. See for instance [Hartle, 1995] and [Rovelli, 2004]. Formulations of this kind are sometimes denoted "generalized quantum mechanics", although they might be called "quantum mechanics" in the same sense in which "classical mechanics" is used to designate formalisms with different degrees of generality, such as Newton's, Lagrange's, Hamilton's or symplectic mechanics.

On the other hand, most of the conventional machinery of perturbative QFT is profoundly incompatible with the general relativistic framework. There are many reasons for this: (i) The conventional formalism of QFT relies on Poincaré invariance. In particular, it relies on the notion of energy and on the existence of the nonvanishing hamiltonian operator that generates unitary time evolution. The vacuum, for instance, is the state that minimizes the energy. But, in a general relativistic theory there is, in general, no global Poincaré invariance, no general notion of energy and no nonvanishing hamiltonian operator. (ii) At the roots of conventional QFT is the physical notion of particle. The theoretical experience with QFT on curved spacetime [Fulling, 1989] and on the relation between acceleration and temperature in QFT [Wald, 1994] indicates that in a generic gravitational situation the notion of particle can be quite delicate. (iii) Consider a conventional renormalized QFT. The physical content of the theory can be expressed in terms of its n-point functions $W(x_1, \ldots, x_n)$. We expect the n-point functions to be invariant under the invariances of the theory. In a general relativistic theory, invariance under an arbitrary coordinate transformation $x \to x' = x'(x)$ implies immediately that the n-point functions must satisfy

$$(15)\quad W(x_1, \ldots, x_n) = W(x'(x_1), \ldots, x'(x_n)).$$

Since any set of n (distict) points $(x_1, \ldots, x_n)$ can be transformed into any other set by a generic coordinate transformation, it follows that W is constant! It does depend on its arguments! Clearly we are in a very different framework from conventional QFT.

There is a possible escape strategy to circumvent these difficulties: write the gravitational field as the sum of two terms, as in equation (1), and assume that spacetime and causal relations are defined by the first term, rather than by the full gravitational field. This escape strategy brings back a background spacetime. A formulation of QG that does *not* take this escape strategy, and thus maintains the full symmetry of GR, is called *background independent.*

The divide

Different research directions are oriented by different evaluations given to the general relativistic spacetime conceptual revolution discussed above in Section 4.1. If this conceptual revolution is taken seriously, and understood as a feature of the world that we have learned, the problem of QG becomes the problem of understanding how to define and interpret a background independent QFT. This point of view orients a large part of the research in loop quantum gravity and similar approaches. Not surprisingly, this line of research is more strongly influenced by the GR research tradition.

On the other hand, if the GR conceptual shift is viewed as accidental, the motivation for developing QG comes more from other open problems such as the problem of the unification (see below). One argument often presented for this point of view is that since QG affects microphysics, we can always choose a scale which is sufficiently small to disregard macroscopic curvature effects and sufficiently large to disregard QG effects. At this scale the world is Lorentz invariant. Therefore in QG we can always assume the existence of an asymptotic Lorentz invariant region. This suggests that we can use techniques associated with asymptotic Lorentz invariance. This line of thinking, predominant in the string community, is more influenced by the particle-physics tradition, which is deeply wedded to Poincaré invariance and which has mostly neglected gravity throughout the twentieth century.

The cultural divide is sometimes very strong, in spite of repeated efforts to fill the gap. Both sides feel that the other side is incapable of appreciating something basic and essential: the structure of QFT as it has been understood in half a century of investigation, for the particle-physics side; the novel physical understanding of space and time that has appeared with GR, for the relativity side. Both sides expect that the other's point of view will turn out, at the end of the day, to be not very relevant. One side because GR is only a low energy limit of a much more complex theory, and thus cannot be taken too seriously as an indication about the deep structure of Nature. The other, because the experience with QFT is on a fixed metric spacetime, and thus is irrelevant in a genuinely background independent context.

4.3 The nature of time

Much has been written about the fact that the main equation of nonperturbative QG, namely the Wheeler-DeWitt equation (5) does not contain the time variable t. This presentation of the "problem of time in QG", however, is misleading, since it confuses the aspect of the problem that is specific to QG and the one which is already present in classical GR. Indeed, classical GR can be entirely formulated in the Hamilton-Jacobi formalism in terms of equation (4), where no time variable appears either.

In the classical general-relativistic context, the notion of time differs strongly from the one used in the special-relativistic context (and even more strongly from the one used in the pre-relativistic context). In the pre-relativistic context, following Newton, we assume that there is a universal physical variable t, measured by clocks, such that all physical phenomena can be described in terms of evolution equations in the independent variable t. In the special-relativistic concept, this notion of time is weakened. Clocks do not measure an universal time variable, but rather a proper time elapsed along inertial trajectories. If we fix a Lorentz frame, however, we can still describe all physical phenomena in terms of evolution equations in the independent variable x^0, even though this description hides the covariance of the system.

In the general relativistic context, we must distinguish two kinds of problems, that are often improperly confused. First, we can consider the problem of the dynamics of matter interacting with a given gravitational field, or, equivalently, on a given spacetime geometry. In this case, the fixed gravitational field still determines the value of the proper time τ elapsed along any (timelike) spacetime trajectory, measured by a clock moving along that trajectory. That is, a given gravitational field determines a local notion of time.

A distinct problem is given by the dynamics of the gravitational field itself, or by the interacting dynamics of gravity and matter. In this case, there is no external time variable that can play the role of observable independent evolution variable. The field equations are written in terms of an evolution parameter, which is the time coordinate x^0, but this coordinate, as explained above in section 4.1, does not correspond to anything observable. In general, the proper time τ along spacetime trajectories also cannot be used as an independent variable, as τ is a complicated non-local function of the gravitational field itself. Therefore, properly speaking, GR does not admit a description as a system evolving in terms of an observable time variable. This is particularly evident in the Hamilton-Jacobi formulation (4) of GR. This does not mean that GR lacks predictivity. Simply put, what GR predicts are relations between partial observables, which cannot in general be represented as dependence of dependent variables on a preferred independent time variable.

To be sure, the ontological status of the time variable t is far from being straightforward in Newtonian physics either. In Newtonian physics we describe the world in terms of physical variables $A(t), B(t), \dots$ evolving in t. One may notice that in a sense we never directly access t, but only physical variables $A, B \dots$, since the clock devices used to measure t are themselves physical systems with an observable time-dependent variable $C(t)$, such as the position of the clock's hand. Therefore, what we actually observe is always the relative evolution of observable variables $A(C), B(C), A(B) \dots$ and never t itself. Newton makes this point clearly in the *Principia*, but also observes that the direct mathematicization of the apparent motions $A(C), B(C), A(B) \dots$ becomes greatly simplified by hypostatizing the existence of t, and expressing all evolution in terms of t. This of course works excellently in the nonrelativistic and nongravitational context. But it is not illogical that Newton's strategy might fail in certain regimes. And in fact it fails in the relativistic gravitational regime, where no universal t can be introduced, and we can only describe the relative dependence of observable quantities. This is what happens in GR.

In a sense, any partial observable variable can be chosen as the independent one in GR. In general, none has the idealized properties assumed by the Newtonian time t, which grows monotonically irrespectively of the state of the system. For instance, in a closed cosmology the volume a of the universe and the proper time t_c since the Big Bang, along a galaxy worldline, are often used as independent variables. But a behaves badly if the Universe begins recontracting, and t_c is only defined in the approximation in which the Universe is assumed to be homogeneous (what is the value of t_c if two galaxies with different proper time from the big bang meet?)

Such a weakening of the notion of time in classical GR is rarely emphasized, because, after all, in classical physics we may disregard the full dynamical structure of the dynamical theory and consider only a single solution of its equations of motion. As mentioned, a single solution of the GR equations of motion determines a spacetime, where a notion of proper time is associated to each timelike worldline. In the quantum context, on the other hand, there is no single spacetime, as there is no trajectory for a quantum particle, and the very concept of time becomes fuzzy.

Attitudes towards the problem of time

Different attitudes can be found in the literature with regard to the problem of time. For technical overviews and references (not completely up to date), see for instance [Isham, 1992; Kuchar, 1992].

As already mentioned, a considerable part of QG research disregards the issue, and maintains that Minkowski space, Poincaré invariance, with its associated notion of time evolution (as a subgroup of the Poincaré group), should not be abandoned in building QG, notwithstanding the features of GR.

Other authors maintain that even if Poincaré invariance is lost in GR and the notion of time becomes more complex, still the idea that the world exists in time, and that its description is the description of systems evolving in time, is a primary notion that we cannot renounce.

Some of these authors have proposed minor modifications of GR, capable of reintroducing a fundamental notion of observable time evolution in the theory. One possibility is to choose a preferred gauge-fixing, in which diffeomorphism invariance is partially broken, and the time coordinate is gauge fixed to be equal to some function of the gravitational field. An example is York time, defined as the trace of the extrinsic curvature of a spacelike surface. Alternatively, the dynamics of GR can be modified, to get a theory with an independent time parameter.

Others accepts in full the challenge presented by GR of trying to conceptualize the world in the absence of a fundamental notion of time and time evolution, as illustrated in the following section.

Physics without space and time?

An illustrative example of how a formulation of mechanics might not use space and time as independent variables is provided by the following proposal (see [Rovelli, 2004], Chapters 4 and 6). Consider a finite spacetime region R bounded by a closed three-dimensional surface Σ. Let (φ, g) represent the value of all fields, including the gravitational field g, on Σ, and let (P_φ, P_g) represent the normal derivative of the fields out of Σ. In principle, all predictions of classical GR can be expressed as constraints on the possible values that the set $(\varphi, g), (P_\varphi, P_g)$ can take. Similarly, in principle all the predictions of QG can be expressed in terms of the probability amplitude $W(\varphi, g)$ of measuring the fields (φ, g). This is a generalization of Feynman's observation that the quantum dynamics of a particle is contained in the propagator $W(x', t'; x, t)$. Diffeomorphism invariance implies that $W(\varphi, g)$ does not depend on the way Σ is imbedded into M. In other words, the entire quantitative spatial and temporal dependence is encoded into the dependence of $W(\varphi, g)$ on the gravitational field g. If, for instance, we identify Σ with the surface of the initial, final and boundary values of a scattering experiment, then it is only the value of the gravitational field on the boundary that determines the time lapses between the initial and final surfaces. Recall indeed that in GR spatial distances and temporal intervals are functions of the gravitational field. $W(\varphi, g)$ can then be used in principle to determine all possible probabilistic predictions regarding the experiment, without using independent spatial or temporal variables.

In order to understand the quantum gravitational field, some of the emphasis on geometry should probably be abandoned. Geometry represents well the classical gravitational field, not quantum spacetime. This is not a betrayal of Einstein's legacy: on the contrary, it is a step in the direction of "relativity" in the precise sense meant by Einstein. The key conceptual difficulty of QG may therefore to find a way to understand the physical world in the absence of the familiar stage of space and time. What might be needed is to free ourselves from the prejudices associated with the habit of thinking of the world as "inhabiting space" and "evolving in time".[2]

Whether it is logically possible to understand the world in the absence of fundamental notions of time and time evolution, and whether this may be consistent with our experience of the world is an open question.

Unitarity

Absence of a fundamental notion of time evolution implies in particular that there is no unitary time evolution in the theory. Absence of unitarity is viewed with great suspicion by many physicists coming from the high energy tradition, where the requirement of unitarity has repeatedly played a major historical role. The argument is often put forward that a probabilistic theory without unitary time evolution is inconsistent. This is not correct, since inconsistency follows from lack of unitarity in the presence of a standard time evolution, and not in the absence of it. If, for instance, we describe the evolution of the universe using the volume of

[2]If we take this extreme attitude, one problem is to recover the macroscopic notion of time evolution and the specific features of the macroscopic time observable, from an atemporal microscopic theory. It is well known that it is surprisingly hard to pin-point with precision what characterizes the time variable in a dynamical system; on the other hand, the thermodynamical and statistical behavior of physical systems is strongly temporally characterized. Accordingly, the hypothesis has been considered [Rovelli, 1993a; Rovelli, 1993b; Connes and Rovelli, 1994] that "temporal flow" is a feature of the world that appears only in the context of a statistical-thermodynamical description. In other words, "time" could be an artifact of our vaste ignorance of the microstate of the world.

the universe a as independent variable, there is no reason to require the probability for the universe to exist to be unit at all a. Indeed, there is a finite probability that the universe reach only a maximum value of a and then re contracts. On the other hand, the consistency of a probabilistic interpretation of QM in a context in which evolution is not expressed in terms of an external variable t is still unclear.

5 RELATION WITH OTHER OPEN PROBLEMS

In the history of physics, often two open problems have found a common solution. For instance, the problem of understanding the nature of light and the problem of unifying the electric and magnetic theory found a common solution in Maxwell theory. Often, however, the hope to solve two problems at once has been disappointed. For instance, in the sixties the hope was strong to find a theory for the strong interaction and at the same time get rid of renormalization theory; but QCD turned out to be a good solution of the first problem without addressing the second. The problem of QG has been suggested to be related to all sorts of open problems in theoretical physics.

5.1 Unification

The current description of the physical world is composed by a number of field theories: on the one hand GR, on the other hand the standard model which in turn is composed of the electroweak theory and quantum chromodynamics; in addition, fermions are present in several multiplets, and there are the Higgs scalars. The theory has more than a dozen elementary constants. In the wake of the successful unifications of electric and magnetic theory, and then of the electromagnetic and the weak interactions theories, research has long aimed to reduce the complexity of the standard model by providing a single coherent theory governed by a smaller number of elementary constants.

Opinions diverge on the relation between this “unification” problem and the problem of QG. *A priori*, there is no strict reason why the quantum properties of gravity should be understood only in conjunction with the other field theories; the quantum properties of electromagnetism, for instance, have been understood in the context of QED without reference to the other interactions, and so have the properties of the strong interactions.

Some arguments have been proposed to support the idea that the two problems must be solved together. I mention three: the first is speculative and I think weack. The second and third are technical and have some weight. First, there is a widespread expectation that a final “Theory Of Everything” should be at hand today. Historically, however, this expectation has been often present in theoretical physics, and so far always erroneously.

The second argument comes from the early history of the attempts at replacing GR with a renormalizable theory: supergravity has shown that the gravitational ultraviolet divergences are suppressed (although, at the end of the day, not cured) by an appropriate coupling between gravity and matter (a fermion field, in the case of supergravity).

The third argument supporting the relation between the two problems is the following. In the standard model, the coupling constants that determine the strength of the electromagnetic, weak and strong interactions depend on the scale of the phenomena considered. At normal scales, they are widely different in size, but they converge at a scale which is quite close to the Planck scale. This suggests that the scale at which unification might take place should be the same as the scale at which quantum gravitational effects become manifest, indicating that the two phenomena are likely to be related.

Concretely, QG is realized in string theory in the context of a tight unification, while loop quantum gravity proposes a solution of the QG problem unrelated to unification.

5.2 Interpretation of quantum mechanics

In spite of its enormous empirical success and its nowadays ubiquitous applications, QM is a theory which is viewed by many as not yet completely understood. The interpretation of the theory is relatively uncontroversial as long as we use it to describe physical systems interacting with an external system (the "observer") whose quantum properties can be disregarded. But a number of difficulties appear as soon as we take the quantum properties of the observer into account. In the physics community, the attitudes to this problem vary widely, ranging from a complete denial that a problem exists to various proposals for modifying QM in order to solve it. But the number of physicists who consider this a genuine open problem has been increasing in the last decade.

Various arguments have been proposed to tie the problem of the interpretation of QM to the QG problem. One is, once more, the expectation that a final theory might be at hand, and the final theory must be entirely self consistent.

Roger Penrose has proposed a specific mechanism via which quantum linearity might be broken by gravity: gravity might the a physical factor inducing a physically realized wave-function collapse [Penrose, 1986]. The proposal is in principle empirically testable.

In the context of Smolin's and Adler's attempts to derive quantum mechanics from the statistical behavior of a statistical dynamics of matrix models [Smolin, 2002; Adler, 2004], the suggestion has been made that one might seek a common origin for both gravitation and quantum field theory at a common deeper level of physical phenomena from which quantum field theory emerges.

Finally, the suggestion has been made that the relational aspect of spatiotemporal structure revealed by GR could be connected with the relational aspect of QM emphasized by the "relational" interpretations of QM [Rovelli, 1996b]. The first is determined by the relation of contiguity between systems; the second by the interaction between systems. But on the one hand locality implies that interaction happens only between contiguous systems, and on the other hand contiguity is only manifest via a physical interaction, suggesting a strict connection between the two relations. These ideas however, have not been developed beyond the stage of suggestions.

5.3 The cosmological constant

An elusive aspect of the current description of the universe is given by the cosmological constant, a constant introduced by Einstein, describing a long range gravitational coupling that can modify gravity at large distances. This constant plays a major role in cosmology, in QFT (where quantum field theoretical effects tend to make it unrealistically large), and recent cosmological observations seem to indicate that its value is very small but not, as previously expected, vanishing.

Although nothing clear has so far appeared in QG research concerning this constant, it must be noted that Raphael Sorkin's QG theory predicted a small value of the constant with the correct of order of magnitude, before its observation, as mentioned in Section 2.1.7.

5.4 Quantum cosmology

"Quantum cosmology" indicates the study of the Universe as a whole as a quantum system. There are two distinct problems that go under this name.

The first is the quantum version of the modelling of the dynamics of our Universe: in particular, the study of the quantum features of the dynamical systems obtained under the drastic simplification that the Universe is homogenous. These classical models, such as the Friedmann-Robertson-Walker model, play an important role in cosmology and are believed to give a good description of the large scale features of our Universe. Their quantization is of interest on several grounds. First, it provides a simplified framework in which many of the conceptual difficulties of QG can be examined and solutions can be tested. Second, they can be used to study what a quantum theory of gravity could us concerning the physics near and at the Big Bang itself, where quantum gravitational effects are expected to dominate.

The study of these models has been started in the sixties by Bryce DeWitt [DeWitt, 1967b; DeWitt, 1967c] and Charles Misner [Misner, 1969] and has seen a great development in the following decades. The limitation of these models, of course, is that they are based on the freezing of all the infinite numbers of degrees of freedom of GR, except for a finite number of them, and therefore they miss the entire field theoretical aspect of the QG problem.

A string cosmology has been developed by Gabriele Veneziano and collaborators, with the hope of finding observational consequences of string theory [Gasperini and Veneziano, 1993]. The application of loop quantum gravity to quantum cosmology ("loop cosmology") has recently led to a model which is finite and well-behaved at the initial singularity [Bojowald, 2001; Bojowald and Morales-Tecotl, 2006].

The second problem that goes under the name of "quantum cosmology" is the conceptual problem of describing a quantum system that forms the entire universe, and therefore for which there is no "external" observer: i.e., the study of quantum mechanics in the case in which the observer is inside the system.

This second problem is very loosely related to the problem of quantum gravity. It is true that it is impossible to be "external" with respect to the gravitational field, but one should not confuse "external" in the spatiotemporal sense with "external" in the dynamical sense. One cannot be "external" with respect to the electromagnetic field either, in the spatiotemporal sense; but we can nevertheless consider an electromagnetic system, viewed as a quantum system, interacting with an external system, viewed as a classical observer. The same can be done for a gravitational system. Therefore nothing *a priori* prevents us from using the standard Copenhagen interpretation of QM (whether or not this is satisfactory) in the context of QG. In other words, the problem of QG and this second problem of quantum cosmology are not necessarily related.

On the other hand, the difficulties raised by considering the observer as part of the system and the difficulties generated in QM by diffeomorphism invariance, in particular the absence of an external time, are of a similar nature, and both question the viability of the Copenhagen interpretation. A general scheme for addressing both kinds of difficulties, and defining a generalized formalism for QM, where there is no external time and no external observer, has been developed by Jim Hartle [Hartle, 1995].

6 CONCLUSION

After 70 years of research, there is no consensus, no established theory, and no QG theory has yet received any direct or indirect experimental support. In the course of 70 years, many ideas have been explored, fashions have come and gone, the discovery of the Holy Grail of QG has been several times announced, only to be later greeted by much scorn.

However, in spite of this, research in QG has not been meandering meaninglessly. On the contrary, a consistent logic has guided the development of the research, from the early formulation of the problem and the research directions in the fifties to nowadays. The implementation of the programs has been laborious, but has been achieved. Difficulties have appeared, and solutions have been proposed, which, after much difficulty, have lead to the realization, at least partial, of the initial hopes.

It was suggested in the early seventies that GR could perhaps be seen as the low energy limit of a theory without uncontrollable divergences; today, 30 years later, such a theory — string theory — is known. In 1957 Charles Misner indicated that in the canonical framework one should be able to compute eigenvalues; and in 1995, 37 years later, eigenvalues were computed — within loop quantum gravity. Much remains to be understood and some of the current developments might lead nowhere. We are not at the end of the road, we are only half-way through the woods. But looking at the entire development of the subject, it is difficult to deny that there has been substantial progress.

The progress cannot be just technical. The search for a quantum theory of gravity raises again old questions such as: What is space? What is time? What is the meaning of "being somewhere"? What is the meaning of "moving"? Is motion to be defined with respect to objects or with respect to space? Can we formulate physics without referring to time or to spacetime? And also: What is causality? What is the role of the observer in physics?

Questions of this kind have played a central role in periods of major advances in physics. For instance, they played a central role for Einstein, Heisenberg, Bohr and their colleagues. But also for Descartes, Galileo, Newton and their contemporaries, and for Faraday, Maxwell and their colleagues. Today some physicists view this manner of posing problems as "too philosophical". Most physicists of the second half of the twentieth century, indeed, have viewed questions of this nature as irrelevant. This view was appropriate for the problems they were facing. When the basics are clear and the issue is problem-solving within a given conceptual scheme, there is no reason to worry about foundations: a pragmatic approach is the most effective one. Today the kind of difficulties that fundamental physics faces has changed. To understand quantum spacetime, physics has to return, once more, to those foundational issues. We have to find new answers to the old foundational questions. The new answers have to take into account what we have learned with QM and GR. The problem of QG will probably not be solved unless these questions are carefully reconsidered.

ACKNOWLEDGEMENTS

Thanks to Christian Wüthrich, Jeremy Butterfield and John Earman for their careful reading of the manuscript and very valuable suggestions.

BIBLIOGRAPHICAL NOTE

For more details on the history of QG see the historical appendix in [Rovelli, 2004]; and, for early history see [Stachel, 1999b; Stachel, 1999a] and [Gorelik, 1992]. For orientation on current research on QG, see the review papers [Horowitz, 2000; Carlip, 2001; Isham, 1991; Rovelli, 1998a]. An interesting panorama of points of view on the problem and on philosophical issues it raises is in the various contributions to the book [Callender and Huggett, 2001]. See also the discussion in [Rovelli, 1997; Rovelli, 2000]. As a general introduction to QG ideas, see the old classic reviews, which are rich in ideas and present different points of view, such as John Wheeler 1967 [Wheeler, 1968], Steven Weinberg 1979 [Weinberg, 1979], Stephen Hawking 1979 and 1980 [Hawking, 1979; Hawking, 1984], Karel Kuchar 1980 [Kuchar, 1984], and Chris Isham's magisterial syntheses [Isham, 1984a; Isham, 1984b; Isham, 1997]. On string theory, classic textbooks are Green, Schwarz and Witten, and

Polchinksi [Green *et al.*, 1987; Polchinski, 1998]. On loop QG, see [Rovelli, 1998b; Rovelli, 2004]. For a discussion of the difficulties of string theory and a comparison of the results of strings and loops, see [Rovelli, 2003], written in the form of a dialogue, and [Smolin, 2003]. Smolin's popular book [Smolin, 2000] provides a readable introduction to QG.

BIBLIOGRAPHY

[Adler, 2004] S. Adler. *Quantum Theory as an Emergent Phenomenon.* Cambridge University Press, 2004.

[Amati *et al.*, 1989] D. Amati, M. Ciafaloni, and G. Veneziano. Can spacetime be probed below the string size? *Physics Letters B*, 216:41–47, 1989.

[Ambjorn *et al.*, 1997] J. Ambjorn, M. Carfora, and A. Marzuoli. *The Geometry of Dynamical Triangulations.* Lecture Notes in Physics. Springer-Verlag, Berlin, 1997.

[Arnowitt *et al.*, 1962] R. Arnowitt, S. Deser, C. Misner. The dynamics of general relativity. In L. Witten (ed.), *Gravitation: An Introduction to Current Research*, page 227. Wiley, New York, 1962.

[Ashtekar, 1986] A. Ashtekar. New variables for classical and quantum gravity. *Physical Review Letters*, 57:2244–2247, 1986.

[Ashtekar, 1987] A. Ashtekar. New Hamiltonian formulation of general relativity. *Physical Review D*, 36:1587–1602, 1987.

[Ashtekar *et al.*, 1998] A. Ashtekar, J. Baez, A. Corichi, and K. Krasnov. Quantum geometry and black hole entropy. *Physical Review Letters*, 80:904–907, 1998. gr-qc/9710007.

[Bardeen *et al.*, 1973] J. Bardeen, B. Carter, and S. Hawking. The four laws of black hole mechanics. *Communication in Mathematical Physics*, 31:161–170, 1973.

[Bekenstein, 1972] J. Bekenstein. Black holes and the second law. *Nuovo Cimento Letters*, 4:737–740, 1972.

[Bekenstein, 1973] J. Bekenstein. Black holes and entropy. *Physical Review*, D7:2333–2346, 1973.

[Bekenstein, 1974] J. Bekenstein. Generalized second law for thermodynamics in black hole physics, 1974. *Physical Review*, D9:3292–3300.

[Belot, 1998] G. Belot. Why general relativity does need an interpretation. *Philosophy of Science*, 63:S80–S88, 1998.

[Bergmann, 1949a] P. Bergmann. Non-linear field theories. *Physical Review*, 75:680–685, 1949a.

[Bergmann, 1949b] P. Bergmann. Non-linear field theories II: Canonical equations and quantization. *Review of Modern Physics*, 21:480–487, 1949b.

[Bergmann, 1958] P. Bergmann. Conservation laws in general relativity as the generators of coordinate transformations. *Physical Review*, 112:287–289, 1958.

[Bergmann, 1961] P. Bergmann. Observables in general relativity. *Review of Modern Physics*, 33:510–514, 1961.

[Bergmann and Komar, 1980] P. Bergmann and A. Komar. The phase space formulation of general relativity and approaches towards quantization. In A. Held (ed.), *General Relativity and Gravitation: One hundred years after the birth of Einstein*, pages 227–254. Plenum Press, New York, 1980.

[Bojowald, 2001] M. Bojowald. Absence of singularity in loop quantum cosmology. *Physical Review Letters*, 86:5227–5230, 2001. gr-qc/0102069.

[Bojowald and Morales-Tecotl, 2006] M. Bojowald, H. Morales-Tecotl. Cosmological applications of loop quantum gravity. gr-qc/0306008. to appear in the proceedings of the Fifth Mexican School (DGFM): *The Early Universe and Observational Cosmology.*, 2006.

[Bronstein, 1936] M. Bronstein. Quantentheories schwacher Gravitationsfelder. *Physikalische Zeitschrift der Sowietunion*, 9:140, 1936.

[Callender and Huggett, 2001] C. Callender and N. Huggett (eds.). *Physics Meets Philosophy at the Planck scale.* Cambridge University Press, Cambridge, 2001.

[Carlip, 2001] S. Carlip. Quantum gravity: A progress report. *Reports on Progress in Physics*, 64:885, 2001. gr-qc/0108040.

[Chamseddine and Connes, 1996] A. Chamseddine and A. Connes. Universal formula for noncommutative geometry actions: Unification of gravity and the standard model. *Physical Review Letters*, 24:4868–4871, 1996.

[Chamseddine and Connes, 1997] A. Chamseddine and A. Connes. The spectral action principle. *Communication in Mathematical Physics*, 186:731–750, 1997.

[Connes, 1994] A. Connes. *Non Commutative Geometry.* Academic Press, New York, 1994.

[Connes and Rovelli, 1994] A. Connes and C. Rovelli. Von Neumann algebra automorphisms and time versus thermodynamics relation in general covariant quantum theories. *Classical and Quantum Gravity*, 11:2899–2917, 1994.

[Deser and van Nieuwenhuizen, 1974a] S. Deser and P. van Nieuwenhuizen. One loop divergences of the quantized Einstein-Maxwell fields. *Physical Review D*, 10:401–410, 1974a.

[Deser and van Nieuwenhuizen, 1974b] S. Deser and P. van Nieuwenhuizen. Nonrenormalizability of the quantized Dirac-Einstein system. *Physical Review*, D10:411–420, 1974b.

[DeWitt, 1964a] B. DeWitt. In Gauthier-Villars (ed.), *Conférence internationale sur les théories relativistes de la gravitation.* Editions Scientifiques de Pologne, Warsaw, 1964a.

[DeWitt, 1964b] B. DeWitt. Theory of radiative corrections for non-Abelian gauge fields. *Physical Review Letters*, 12:742–746, 1964b.

[DeWitt, 1965] B. DeWitt. *Dynamical Theory of Groups and Fields".* Wiley, New York, 1965.

[DeWitt, 1967a] B. DeWitt. Quantum theory of gravity. I. The canonical theory. *Physical Review*, 160:1113–1148, 1967a.

[DeWitt, 1967b] B. DeWitt. Quantum theory of gravity. II. The manifestly covariant theory. *Physical Review*, 162:1195–1239, 1967b.

[DeWitt, 1967c] B. DeWitt. Quantum theory of gravity. III. Applications of the covariant theory. *Physical Review*, 162:1239–1256, 1967c.

[Dirac, 1950] P. Dirac. Generalized Hamiltonian dynamics. *Canadian Journal of Physics*, 2:129–148, 1950.

[Dirac, 1958] P. Dirac. The theory of gravitation in Hamiltonian form. *Proceedings of the Royalyal Society of London*, A246:333–343, 1958.

[Dirac, 1959] P. Dirac. Fixation of coordinates in the Hamiltonian theory of gravitation. *Physical Review*, 114:924–930, 1959.

[Dirac, 1964] P. Dirac. *Lectures on Quantum Mechanics.* Belfer Graduate School of Science, Yeshiva University, New York, 1964.

[Doplicher, 1996] S. Doplicher. Quantum spacetime. New problems in the general theory of fields and particles, part II. *Annales de l'Institut Henri Poincare*, 64:543–553, 1996.

[Doplicher *et al.*, 1994] S. Doplicher, K. Fredenhagen, J. Roberts. Spacetime quantization induced by classical gravity. *Physics Letters B*, 331:39–44, 1994.

[Doplicher *et al.*, 1995] S. Doplicher, K. Fredenhagen, J. Roberts. The quantum structure of spacetime at the Planck scale and quantum fields. *Communications in Mathematical Physics*, 172:187–220, 1995.

[Earman,] J. Earman. Private communication.

[Earman, 1987] J Earman, J. Norton. What price spacetime substantivalism? The hole story. *British Journal for the Philosophy of Science*, 38:515–525, 1987.

[Earman, 1989] J. Earman. *World Enough and Space-time: Absolute versus Relational Theories of Spacetime.* MIT Press, Cambridge, 1989.

[Earman, 2001] J. Earman and G. Belot. Pre-Socratic quantum gravity. In C. Callander (ed.), *Physics meets philosphy at the Planck scale.* Cambridge University Press, Cambridge, 2001.

[Einstein, 1916] A. Einstein. Naeherungsweise integration der Feldgleichungen der gravitation. *Preussische Akademie der Wissenschaften (Berlin) Sitzungsberichte*, pages 688–696, 1916.

[Einstein and Grossmann, 1914] A. Einstein and M. Grossmann. Entwurf einer verallgemeinerten Relativitätstheorie und einer Theorie der Gravitation. *Zeitschrift für Mathematik und Physik*, 62:225–261, 1914.

[Faddeev and Popov, 1967] L. Faddeev and V. Popov. Feynman diagrams for the Yang-Mills field. *Physics Letter B*, 25:29–30, 1967.

[Feynman, 1963] R. Feynman. Quantum theory of gravitation. *Acta Physical Polonica*, 24:697–722, 1963.

[Fierz, 1939] M. Fierz. Ber die relativistische Theorie kräfterfreier Teilchen mit beliebigem Spin. *Helvetica Physica Acta*, 12:3–37, 1939.

[Finkelstein, 1997] D. Finkelstein. *Quantum Relativity.* Springer, Berlin, 1997.

[Freedman *et al.*, 1976] D. Freedman, S. Ferrara, and P. van Nieuwenhuizen. Progress toward a theory of supergravity. *Physical Review*, D13:3214–3218, 1976.

[Fulling, 1989] S. Fulling. *Aspects of Quantum Field Theory in Curved Spacetime*. Cambridge University Press, Cambridge, 1989.

[Gasperini and Veneziano, 1993] M. Gasperini and G. Veneziano. Pre-bigbang in string cosmology. *Astroparticle Physics*, 1:317–339, 1993.

[Gorelik, 1992] G. Gorelik. First steps of quantum gravity and the Planck values. In J. Eisenstaedt and A. Kox (eds.), *Studies in the history of general relativity*, volume 3 of *Einstein Studies*, pages 364–379. Birkhaeuser, Boston, 1992.

[Gorelik and Frenkel, 1994] G. Gorelik and V. Frenkel. *Matvei Petrovic Bronstein and the Soviet Theoretical Physics in the Thirties*. Birkhauser Verlag, Boston, 1994.

[Goroff and Sagnotti, 1985] M. H. Goroff, A. Sagnotti. Quantum gravity at two loops. *Physics Letter B*, 160:81–86, 1985.

[Goroff and Sagnotti, 1986] M. H. Goroff, A. Sagnotti. The ultraviolet behavour of Einstein gravity. *Nuclear Physics B*, 266:709–736, 1986.

[Green and Schwarz, 1984] M. Green and J. Schwarz. Anomaly cancellation in supersymmetric d=10 gauge theory requires SO(32). *Physics Letters B*, 149:117–122, 1984.

[Green *et al.*, 1987] M. Green, J. Schwarz, and E. Witten. *Superstring Theory*. Cambridge University Press, Cambridge, 1987.

[Gupta, 1952] S. Gupta. Quantization of Einstein's gravitational field: General treatment. *Proceedings Physical Society A*, 65:608–619, 1952.

[Hall and Hall, 1962] A. Hall and M. Hall (eds.). *Unpublished papers of Isaac Newton*. Cambridge University Press, Cambridge, 1962.

[Hartle, 1995] J. Hartle. Spacetime quantum mechanics and the quantum mechanics of spacetime. In B Julia, J. Z.-J. (eds.), *Proceedings on the 1992 Les Houches School, Gravitation and Quantisation*, page 285. Elsevier Science, Paris, 1995.

[Hartle and Hawking, 1983] J. Hartle and S. Hawking. Wave function of the universe. *Physical Review*, D28:2960–2975, 1983.

[Hawking, 1974] S. Hawking. Black hole explosions? *Nature*, 248:30–31, 1974.

[Hawking, 1975] S. Hawking. Particle creation by black holes. *Communication in Mathematical Physics*, 43:199–220, 1975.

[Hawking, 1979] S. Hawking. The path-integral approach to quantum gravity. In S. Hawking and W. Israel (eds.), *General Relativity: An Einstein Centenary Survey*. Cambridge University Press, Cambridge, 1979.

[Hawking, 1984] S. Hawking. Quantum cosmology. In B. DeWitt and R. Stora (eds.), *Relativity, Groups and Topology, Les Houches Session XL*. North Holland, Amsterdam, 1984.

[Heisenberg, 1938] W. Heisenberg. Die Grenzen der Anwendbarkeit der bisherigen Quantentheorie. *Zeitschrift für Physik*, 110:251–266, 1938.

[Horowitz, 2000] G. Horowitz. Quantum gravity at the turn of the millenium. Plenary talk at the Marcell Grossmann conference, Rome, 2000. gr-qc/0011089.

[Howard and Stachel, 1989] D. Howard and J. Stachel (eds.). *Einstein and the History of General Relativity*, volume I of *Einstein Studies*. Birkhauser, Boston, 1989.

[Isham, 1984a] C. Isham. In B. DeWitt and R. Stora (eds.), *Relativity Groups and Topology. Les Houches 1983*. North Holland, Amsterdam, 1984a.

[Isham, 1984b] C. Isham. Quantum gravity: an overview. In *Oxford 1980, Proceedings, Quantum Gravity 2*. Oxford University Press, Oxford, 1984b.

[Isham, 1992] C. Isham. Canonical quantum gravity and the problem of time. Lectures presented at the NATO Advanced Institute: "Recent problems in Mathematical Physics, Salamanca June 15", 1992.

[Isham, 1994] C. Isham. Quantum logic and the histories approach to quantum theory. *Journal of Mathematical Physics*, 35:2157–2185, 1994. gr-qc/9308006.

[Isham, 1997] C. Isham. Structural problems facing quantum gravity theory. In M. Francaviglia, G. Longhi, L. Lusanna, and E. Sorace (eds.), *Proceedings of the 14th International Conference on General Relativity and Gravitation*, pages 167–209. World Scientific, Singapore, 1997.

[Isham, 1991] C. J. Isham. Conceptual and geometrical problems in quantum gravity. In H. Mitter and H. Gausterer (eds.), *Recent Aspects of Quantum Fields: Proceedings of the Xxx Int. Universitatswochen Fur Kernphysik, Schladming, Austria February and March 1991 (Lectur)*, page 123. Springer-Verlag, Berlin, 1991.

[Jacobson, 1995] T. Jacobson. Thermodynamics of spacetime: The Einstein equation of state. *Physical Review Letters*, 75:1260–1263, 1995.

[Jacobson and Smolin, 1988] T. Jacobson and L. Smolin. Nonperturbative quantum geometries. *Nuclear Physics B*, 299:295–345, 1988.

[Klein, 1927] O. Klein. Zur fünfdimensionalen Darstellung der Relativitaetstheorie. *Zeitschrift für Physik"*, 46:188–208, 1927.

[Krasnov, 1997] K. Krasnov. Geometrical entropy from loop quantum gravity. *Physical Review D*, 55:3505–3513, 1997.

[Kretschmann, 1917] E. Kretschmann. Über den physikalischen Sinn der Relativitätspostulate. *Annals of Physics (Leipzig)*, 53:575–614, 1917.

[Kuchar, 1984] K. Kuchar. Canonical methods of quantization. In *Oxford 1980, Proceedings, Quantum Gravity 2*. Oxford University Press, Oxford, 1984.

[Kuchar, 1992] K. Kuchar. Time and interpretations of quantum gravity. In G. Kunstatter, D. Vincent, and J. Williams (eds.), *Proceedings of the 4th Canadian conference on General Relativity and Relativistic Astrophysics*, Singapore. World Scientific, 1992.

[Misner, 1957] C. Misner. Feynman quantization of general relativity. *Review of Modern Physics*, 29:497–509, 1957.

[Misner, 1969] C. Misner. Quantum cosmology. I. *Physical Review*, 186:1319–1327, 1969.

[Newton, 1962] I. Newton. De gravitatione et aequipondio fluidorum. Translation in [Hall and Hall, 1962], 1962.

[Norton, 1984] J. Norton. How Einstein found his field equations: 1912-1915. *Historical Studies in the Physical Sciences*, 14:253–316, 1984. Reprinted in [Howard and Stachel, 1989, 101–159].

[Pauli and Fierz, 1939] W. Pauli and M. Fierz. On relativistic field equations of particles with arbitrary spin in an electromagnetic field. *Helvetica Physica Acta*, 12:297–300, 1939.

[Pauri and Vallisneri, 2002] M. Pauri and M. Vallisneri. Ephemeral point-events: is there a last remnant of physical objectivity? *Dialogos*, 79:263–303, 2002.

[Penrose, 1967] R. Penrose. Twistor algebra. *Journal of Mathematical Physics*, 8:345–366, 1967.

[Penrose, 1986] R. Penrose. Gravity and state vector reduction. In R. Penrose and C. Isham (eds.), *Quantum Concepts in Space and Time*, page 129. Clarendon Press, Oxford, 1986.

[Peres, 1962] A. Peres. On Cauchy's problem in general relativity. *Nuovo Cimento*, 26:53–61, 1962.

[Perez, 2003] A. Perez. Spin foam models for quantum gravity. *Classical and Quantum Gravity*, 20:R43–R104, 2003. gr-qc/0301113.

[Polchinski, 1995] J. Polchinski. Dirichlet branes and Ramon-Ramon charges. *Physical Review Letters*, 75:4724–4727, 1995.

[Polchinski, 1998] J. Polchinski. *String Theory*. Cambridge University Press, Cambridge, 1998.

[Rosenfeld, 1930a] L. Rosenfeld. Über die Gravitationswirkungen des Lichtes. *Zeitschrift für Physik*, 65:589–599, 1930a.

[Rosenfeld, 1930b] L. Rosenfeld. Zur quantelung der Wellenfelder. *Annalen der Physik*, 5:113–152, 1930b.

[Rovelli, 1993a] C. Rovelli. Statistical mechanics of gravity and thermodynamical origin of time. *Classical and Quantum Gravity*, 10:1549–1566, 1993a.

[Rovelli, 1993b] C. Rovelli. The statistical state of the universe. *Classical and Quantum Gravity*, 10:1567–1578, 1993b.

[Rovelli, 1996a] C. Rovelli. Black hole entropy from loop quantum gravity. *Physical Review Letters*, 14:3288–3291, 1996a.

[Rovelli, 1996b] C. Rovelli. Relational quantum mechanics. *International Journal of Theoretical Physics*, 35:1637–1678, 1996b.

[Rovelli, 1997] C. Rovelli. Halfway through the woods. In J. Earman and J. Norton (eds.), *The Cosmos of Science*. University of Pittsburgh Press and Universitäts Verlag-Konstanz, 1997.

[Rovelli, 1998a] C. Rovelli. Strings, loops and the others: a critical survey on the present approaches to quantum gravity. In N. Dadhich and J. Narlikar (eds.), *Gravitation and Relativity: At the turn of the Millenium (15th International Conference on General Relativity and Gravitation)*, pages 281–331. Inter-University centre for Astronomy and Astrophysics, Pune, 1998a. gr-qc/9803024.

[Rovelli, 1998b] C. Rovelli. Loop Quantum Gravity. *Living Reviews in Relativity*, 1998b, www.livingreviews.org/Articles/Volume1/1998-1rovelli.

[Rovelli, 2000] C. Rovelli. The century of the incomplete revolution: searching for general relativistic quantum field theory. *Journal of Mathematical Physics*, 41:3776–3800, 2000. hep-th/9910131.

[Rovelli, 2003] C. Rovelli. A dialog on quantum gravity. *International Journal of Modern Physics D*, 12:1509–1528, 2003. hep-th/0310077.

[Rovelli, 2004] C. Rovelli. *Quantum Gravity.* Cambridge University Press, Cambridge, 2004.

[Rovelli and Smolin, 1988] C. Rovelli and L. Smolin. Knot theory and quantum gravity. *Physical Review Letters*, 61:1155, 1988.

[Rovelli and Smolin, 1990] C. Rovelli and L. Smolin. Loop space representation for quantum general relativity. *Nuclear Physics B*, 331:80, 1990.

[Rovelli and Smolin, 1995] C. Rovelli and L. Smolin. Discreteness of Area and Volume in Quantum Gravity. *Nuclear Physics B*, 442:593–622, 1995.

[Rovelli and Speziale, 2003] C. Rovelli and S. Speziale. Reconcile Planck-scale discreteness and the Lorentz-Fitzgerald contraction. *Physical Review D*, 67:064019, 2003.

[Smolin, 2000] L. Smolin. *Three Roads to Quantum Gravity.* Oxford University Press, Oxford, 2000.

[Smolin, 2002] L. Smolin. Matrix models as non-local hidden variables theories, 2002. hep-th/0201031.

[Smolin, 2003] L. Smolin. How far are we from the quantum theory of gravity?, 2003 hep-th/0303185.

[Sorkin, 1983] R. Sorkin. Posets as lattice topologies. In B. Bertotti, F. de Felice, A. Pascolini (eds.), *General Relativity and Gravitation: Proceedings of the GR10 Conference*, volume I, page 635. Consiglio Nazionale Delle Ricerche, Rome, 1983.

[Sorkin, 2002] R. Sorkin. An example relevant to the Kretschmann-Einstein debate. *Modern Physics Letters A*, 17:695–700, 2002.

[Stachel, 1999a] J. Stachel. Early history of quantum gravity. In B. Iyer and B. Bhawal (eds.), *Black Holes, Gravitational radiation and the Universe*. Kluwer Academic Publisher, Netherlands, 1999a.

[Stachel, 1999b] J. Stachel. Early history of quantum gravity (1916-1940), 1999b. Presented at the HGR5, Notre Dame, July 1999.

[Stelle, 1977] K. Stelle. Renormalization of higher derivatives quantum gravity. *Physical Review*, D16:953–969, 1977.

[Strominger and Vafa, 1996] A. Strominger and G. Vafa. Microscopic origin of the Bekenstein-Hawking entropy. *Physics Letter*, B379:99–104, 1996.

['t Hooft, 1973] G. 't Hooft. An algorithm for the poles at dimension 4 in the dimensional regularization procedure. *Nuclear Physics B*, 62:444–460, 1973.

['t Hooft, 1996] G. 't Hooft. The scattering matrix approach for the quantum black hole, an overview. *International Journal of Modern Physics A*, 11:4623–4688, 1996. gr-qc/9607022.

['t Hooft and Veltman, 1974] G. 't Hooft and M. Veltman. One loop divergencies in the theory of gravitation. *Annales de l'Institut Henri Poincare*, 20:69–74, 1974.

[Unruh, 1976] W. Unruh. Notes on black hole evaporation. *Physical Review*, D14:870–892, 1976.

[van Nieuwenhuizen, 1981] P. van Nieuwenhuizen. Supergravity. *Physics Reports*, 68:189–398, 1981.

[Wald, 1994] R. Wald. *Quantum Field Theory on Curved Spacetime and Black Hole Thermodynamics.* University of Chicago Press, Chicago, 1994.

[Weinberg, 1979] S. Weinberg. Ultraviolet divergences in quantum theories of gravitation. In S. Hawking and W. Israel (eds.), *General Relativity: An Einstein Centenary Survey.* Cambridge University Press, Cambridge, 1979.

[Wheeler, 1968] J. Wheeler. Superspace and the nature of quantum geometrodynamics. In C. DeWitt and J. Wheeler (eds.), *Batelle Rencontres, 1967*, volume 242 of *Lectures in Mathematics and Physics*. Benjamin, New York, 1968.

[Williams, 1997] R. Williams. Recent progress in Regge calculus. *Nuclear Physics Proceedings Suppl*, 57:73–81, 1997. gr-qc/9702006.

[Williams and Tuckey, 1992] R. Williams and P. Tuckey. Regge calculus: A bibliography and brief review. *Classical and Quantum Gravity*, 9:1409–1422, 1992.

This page intentionally left blank

SYMMETRIES AND INVARIANCES IN CLASSICAL PHYSICS

Katherine Brading and Elena Castellani

1 INTRODUCTION

The term 'symmetry' comes with a variety of ancient connotations, including beauty, harmony, correspondence between parts, balance, equality, proportion, and regularity. These senses of the term are clearly related to one another; the concept of symmetry used in modern physics arose out of this family of ideas. We are familiar with the approximate symmetries of physical objects that we find around us — the bilateral symmetry of the human face, the rotational symmetry of a snowflake turned through 60^o, and so forth. We may define a symmetry of a given geometric figure as the invariance of that figure when equal component parts are exchanged under a specified operation (such as rotation). The development of the algebraic concept of a group, in the nineteenth century, allowed a generalization and refinement of this idea; a precise mathematical notion of symmetry emerged which was applicable not just to physical objects and geometrical figures, but also to mathematical equations — and thus, to what is of particular interest to us, the laws of physics expressed as mathematical equations. The group theoretical notion of symmetry is the notion of invariance under a specified group of transformations. 'Invariance' is a mathematical term: something is invariant when it is left unaltered by a given transformation. This mathematical notion is used to express the notion of physical symmetry that we are interested in, i.e. invariance under a group of transformations. This is the concept of symmetry that has proved so successful in modern science, and the one that will concern us in what follows.

We begin in Section 2 with the distinction between symmetries of objects and of laws, and that between symmetry principles and symmetry arguments. This section includes a discussion of Curie's principle. Section 3 discusses the important connection between symmetries, as studied in physics, and the mathematical techniques of group theory. We offer a brief history of how group theory was applied first to geometry and then to physics in the course of the nineteenth century, preluding to the central importance acquired by group theoretical techniques in contemporary physics. With these considerations in mind, Section 4 offers an account of what is meant by symmetry in physics, and a taxonomy of the different

Handbook of the Philosophy of Science. Philosophy of Physics
Volume editors: Jeremy Butterfield and John Earman
General editors: Dov M. Gabbay, Paul Thagard and John Woods
© 2007 Elsevier B.V. All rights reserved.

types of symmetry that are found within physics. In Section 5 we discuss some applications of symmetries in classical physics, beginning with transformation theory in classical mechanics, and then turning to Einstein's Special and General Theories of Relativity (see Section 6). We focus on the roles and meaning of symmetries in these theories, and this leads into the discussion of Noether's theorems in Section 7. Finally, in Section 8, we offer some concluding remarks concerning the place, role and interpretation of symmetries in classical physics. Note that our emphasis is resolutely on the classical. For the power and significance of symmetry in quantum physics, we refer the reader to other chapters of this volume, such as Dickson (ch. 4, Section 3.3), Landsman (ch. 5, Section 4.1), t'Hooft (ch. 7) and Halvorson (ch. 8, Section 5.2).[1]

2 SYMMETRIES OF OBJECTS AND OF LAWS

That we must distinguish between symmetries of objects versus symmetries of laws can be seen as follows. It is one thing to ask about the geometric symmetries of certain objects — such as the 60^o rotational symmetry of a snowflake and the approximate bilateral symmetry of the human face mentioned above — and the *a*symmetries of objects — such as the failure of a chair to be rotationally symmetric. It is another thing to ask about the symmetries of the laws governing the time-evolution of those objects: we can apply the laws of mechanics to the evolution of our chair, considered as an isolated system, and these laws are rotationally invariant (they do not pick out a preferred orientation in space) even though the chair itself is not. Re-phrasing the same point, we should distinguish between symmetries of *states* or *solutions*, versus symmetries of laws. Having distinguished these two types of symmetry we can, of course, go on to ask about the relationship between them: see, for example, current discussions of Curie's principle, referred to in Section 2.2, below.

2.1 Symmetry principles and symmetry arguments

It is also important to distinguish between *symmetry principles* and *symmetry arguments*. The application of symmetry principles to laws was of central importance to physics in the twentieth century, as we shall see below in the context of Eintein's Special and General Theories of Relativity. Requiring that the laws — whatever their precise form might be — satisfy certain symmetry properties, became a central methodological tool of theoretical physicists in the process of arriving at the detailed form of various laws.

Symmetry arguments, on the other hand, involve drawing specific consequences with regard to particular phenomena on the basis of their symmetry properties. This type of use of symmetry has a long history; examples include Anaximander's argument for the immobility of the Earth, Archimedes's equilibrium law for the

[1]Further discussion can be found in Brading and Castellani [2003].

balance, and the case of *Buridan's ass*.[2] In each case the associated argument can be understood as an example of the application of the Leibnizean Principle of Sufficient Reason (PSR): if there is no sufficient reason for one thing to happen instead of another, then nothing happens (i.e. the initial situation does not change). There is something more that the above cases have in common: in each of them PSR is applied on the grounds that the initial situation has a certain *symmetry*.[3] The symmetry of the initial situation implies the complete equivalence between the offered alternatives. If the alternatives are completely equivalent, then there is no sufficient reason for choosing between them and the initial situation remains unchanged. Arguments of this kind most frequently take the following form: a situation with a certain symmetry evolves in such a way that, in the absence of an asymmetric cause, the initial symmetry is preserved. In other words, a breaking of the initial symmetry cannot happen without a reason: an asymmetry cannot originate spontaneously. This style of argumentation is also to be found in recent discussions of 'Curie's principle', the principle to which we now turn.

2.2 Curie's principle

Pierre Curie (1859-1906) was led to reflect on the question of the relationship between physical properties and symmetry properties of a physical system by his studies on the thermal, electric and magnetic properties of crystals, since these properties were directly related to the structure, and hence the symmetry, of the crystals studied. More precisely, the question he addressed was the following: in a given physical medium (for example, a crystalline medium) having specified symmetry properties, which physical phenomena (for example, which electric and magnetic phenomena) are allowed to happen? His conclusions, systematically presented in his 1894 work 'Sur la symétrie dans les phénomènes physiques', can be summarized as follows:[4]

(a_1) When certain causes produce certain effects, the symmetry elements of the causes must be found in their effects.

(a_2) When certain effects show a certain dissymmetry, this dissymmetry must be found in the causes which gave rise to them.[5]

(a_3) In practice, the converses of these two propositions are not true, i.e., the effects can be more symmetric than their causes.

---

[2] For a discussion of these examples, see [Brading and Castellani, 2003, ch. 1, Section 2.2]).

[3] In the first case rotational symmetry, in the second and third bilateral symmetry.

[4] For an English translation of Curie's paper, see [Curie, 1981]; some aspects of the translation are misleading.

[5] Curie uses the term *dissymmetry* in his paper, as was current at his time. The sense is the same of that of *symmetry breaking* in modern terminology, which is today often identified with the sense of *asymmetry*. To be more precise one should distinguish between the result of a symmetry-breaking process (*broken symmetry*), the absence of *one* of the possible symmetries compatible with the situation considered (*dissymmetry*, as it was called in the nineteenth century literature, notably by Louis Pasteur in his works on molecular dissymmetry), and the absence of *all* the possible symmetries compatible with the situation considered (*asymmetry*).

(b) A phenomenon may exist in a medium having the same characteristic symmetry or the symmetry of a subgroup of its characteristic symmetry. In other words, certain elements of symmetry can coexist with certain phenomena, but they are not necessary. What is necessary, is that certain elements of symmetry do not exist. Dissymmetry is what creates the phenomenon.

Conclusion (a_1) is what is usually called *Curie's principle* in the literature. Conclusion (a_2) is logically equivalent to (a_1); the claim is that symmetries are necessarily transferred from cause to effect, while dissymmetries are not. Conclusion (a_3) clarifies this claim, emphasizing that since dissymmetries need not be transferred from cause to effect, the effect may be more symmetric than the cause.[6] Conclusion (b) invokes a distinction found in all of Curie's examples, between the 'medium' and the 'phenomena'. We have a medium with known symmetry properties, and Curie's principle concerns the relationship between the phenomena that can occur in the medium and the symmetry properties — or rather, '*dis*symmetry' properties — of the medium. Conclusion (b) shows that Curie recognized the important function played by the concept of dissymmetry — of *broken symmetries* in current terminology — in physics.

In order for Curie's principle to be applicable, various conditions need to be satisfied: the cause and effect must be well-defined, the causal connection between them must hold good, and the symmetries of both the cause and the effect must also be well-defined (this involves both the physical and the geometrical properties of the physical systems considered). Curie's principle then furnishes a necessary condition for given phenomena to happen: only those phenomena can happen that are compatible with the symmetry conditions stated by the principle. Curie's principle has thus an important methodological function: on the one hand, it furnishes a kind of selection rule (given an initial situation with a specified symmetry, only certain phenomena are allowed to happen); on the other hand, it offers a falsification criterion for physical theories (a violation of Curie's principle may indicate that something is wrong in the physical description).

Such applications of Curie's principle depend, of course, on our accepting its truth, and this is something that has been questioned in the literature, especially in relation to spontaneous symmetry breaking. Different proposals have been offered for justifying the principle. Curie himself seems to have regarded it as a form of causality principle, and the question in the recent literature has been whether the principle can be demonstrated from premises that include a definition of "cause" and "effect". In this direction it has become current of late to understand the principle as following from the invariance properties of *deterministic* physical laws. The seminal paper for this approach is [Chalmers, 1970], which introduces the formulation of Curie's principle in terms of the relationship between

[6]Note that for some authors conclusion (b) is a principle on its own. Radicati (1987) goes further, describing conclusions (a_1), (a_2) and (b) as three different principles: Curie's first, second and third principle, respectively.

the symmetries of earlier and later states of a system, and the laws connecting these states. This "received view" can be criticized for offering a reformulation that is significantly different from Curie's intentions (so that the label 'Curie's principle' is a misnomer), and for resting on an assumption that may undermine the interest and importance of the view, as we discuss in the following brief remarks.[7]

The received view, by concerning itself with temporally ordered cause and effect pairs (or states of systems), offers a diachronic or dynamic analysis. In fact, Curie himself focusses on synchronic or static situations, concerning the compatibility of different phenomena occurring at the same time, rather than the evolution of one state of a system into another state. In other words, the 'cause—effect' terminology used by Curie is *not* intended to indicate a temporal ordering of phenomena being considered. This is clear from his examples, and also from the fact that discussion of the laws — so central to the diachronic version — is absent from Curie's own analysis. That the diachronic version has come to have the label 'Curie's principle' therefore misrepresents Curie's original principle and his discussion of that principle.

Is the diachronic version interesting and important, nevertheless? The account can be understood as an application of PSR in which we pay careful attention to whether the laws provide a "sufficient reason" for a symmetry to be broken as a system evolves from its initial to final state by means of those laws. The reformulation of the diachronic version by Earman [2004] has the strong merit of being precise, and thereby enabling a proof that if the initial state possesses a given symmetry, and the laws deterministically preserve that symmetry, then the final state will also possess that symmetry. However, things are not so simple as they might seem because the proof takes a state with a *given* symmetry. Specifying the symmetries of a state requires, in general, recourse to a background structure — such as space or spacetime, or the space of solutions. In some cases, the required structure may seem trivial or minimal, but nevertheless the dynamics of the system will not be independent of this structure (consider the examples of the spatial or spatiotemporal structure or, more strongly still, the space of solutions). This has the consequence that, in general, the structures on which the symmetries of a state and the symmetries of the dynamics depend are not independent of one another, and any appearance to the contrary in the "proof" needs to be handled with caution. Indeed, we think that answering the question of whether the diachronic version is interesting and important depends in part upon investigating this lack of independence and the role it plays in the proof, something which has yet to be provided in the literature on the diachronic version of 'Curie's principle'.

Both Curie's original version of his principle, and the diachronic version, begin with the symmetries of states of physical systems. In contemporary physics, focus has shifted to symmetries of laws, and the significant connection between symmetries of physical systems and symmetries of laws has to do not with symmetries

[7]For detailed discussion see [Brading and Castellani, 2006]. The "received view" that we attribute first to Chalmers is developed in [Ismael, 1997] and [Earman, 2004]. See also [Earman, this vol., ch. 14, Section 2.3].

of *states* of those systems, but with symmetries of *ensembles of solutions*.[8] The symmetries of a dynamical equation are not, in general, the symmetries of the individual solutions (let alone states), but rather the symmetries of the whole set of solutions, in the sense that a symmetry of a dynamical equation transforms a given solution into another solution. Considering this relationship between laws and solutions leads to an alternative version of Curie's principle, which we propose here.[9] As with the diachronic version of Curie's principle, our proposal departs from Curie's original proposal, but our contention is that it remains true to the main motivation behind Curie's original investigation. In this version we seek to unite two things:

1. We understand Curie's motivating question to be 'which phenomena are physically possible?', and his suggestion to be that we can use symmetries as a guide towards answering this question; and

2. We go beyond Curie in making use of symmetries of laws, something about which he said nothing, but which has become a central concern in contemporary physics.

Combining these two ingredients, a "modern" version of Curie's principle would then simply state that *the symmetries of the law (equation) are to be found in the ensemble of its solutions.* This version expresses Curie's basic idea — that "symmetry does not get lost (without a reason)" — in virtue of the fact that the symmetry of the law is to be found in the ensemble of solutions. The fact that this is how we define the relationship between symmetries and laws does not render it empty of significance with respect to Curie's motivating question. On the contrary, the point is that we can *use* the symmetries of the law as a guide to finding solutions, i.e. to determining which phenomena are physically possible, when not all the solutions are known. We can ask, following Curie, 'What phenomena are possible?', and we can use the connection between the symmetries of the law and the symmetries of the ensemble of solutions as a guide to finding the physically possible phenomena. Thus, what is on the one hand a definitional statement (that the symmetries of the law (equation) are to be found in the ensemble of the solutions) comes on the other hand to have epistemic bite when we don't know all the solutions. This, we believe, is true to Curie's motivating question, as expressed in item (1), above.

[8]By "solution" here we mean a temporally extended history of a system, the "state" of a system being a "solution at an instant".

[9]Notice that this version does not involve the temporal evolution from cause to effect (as in the diachronic version), nor is it restricted to a state of a system at a given instant or during a certain temporal period (as in the sychronic version); rather, it concerns the structure of an ensemble of solutions, considered as a whole.

3 SYMMETRY AND GROUP THEORY: EARLY HISTORY

Group theory is the powerful mathematical tool by means of which the symmetry properties of theories are studied. In this section, we begin with the definition of a group, and outline the origins of this notion in the mathematics of algebraic equations. We then turn our attention to the manner in which group theory was applied first to geometry and then to physics in the course of the nineteenth century.

3.1 The introduction of the group concept and the first developments of group theory

A *group* is a family $\mathcal{G}$ of elements $g_1, g_2, g_3...$ for which there is defined a multiplication that assigns to every two elements g_i and g_j of the group a third element (their product) $g_k = g_i g_j \in \mathcal{G}$, in such a way that the following requirements hold:[10]

- $(g_i g_j)g_k = g_i(g_j g_k)$ for all $g_i, g_j, g_k \in \mathcal{G}$ (associativity of the product);
- there exists an *identity element* $e \in \mathcal{G}$ such that $g_i e = e g_i$ for all $g_i \in \mathcal{G}$;
- for all $g_i \in \mathcal{G}$ there exists an *inverse element* $g_i^{-1} \in \mathcal{G}$ such that $g_i g_i^{-1} = g_i^{-1} g_i = e$.

The concept of a group was introduced by Évariste Galois in the short time he was able to contribute to mathematics (born in 1811, he died as a result of a duel in 1832) in connection with the question of the resolution of equations by radicals.[11] The resolvent formulas for cubic and quartic equations were found by the mathematicians of the Renaissance,[12] while the existence of a formula for solving the general equations of the fifth and higher degrees by radicals remained an open question for a long time, stimulating developments in algebra. In particular, the studies of the second half of the eighteenth century focussed on the role played in the solution of equations by functions invariant under permutation of the roots, so giving rise to the theory of permutations. In J. L. Lagrange's *Réflexions sur la résolution algébrique des équations*, the most influential text on the subject, some fundamental results of permutation theory were obtained.[13] Lagrange's text

[10]The concept of a group can be weakened by relaxing these conditions (for example, dropping the inverse requirement leads to the concept of a monoid, and retaining only associativity leads to the concept of a semigroup). The question that then arises is whether the full group structure, or some weaker structure, is related to the symmetry properties of a given theory.

[11]That is, in terms of a finite number of algebraic operations — addition, subtraction, multiplication, division, raising to a power and extracting roots — on the coefficients of the equations

[12]The resolvent formula for a quadratic equation was known since Babylonian times. A historical survey on the question of the existence of resolvent formulas for algebraic equations is in [Yaglom, 1988, 3 f.].

[13]Among other results, the so-called Lagrange's theorem which states — in modern terminology — that the order of a subgroup of a finite group is a divisor of the order of the group.

served as a basis for successive algebraic developments, from P. Ruffini's first proof in 1799 of the impossibility of solving nth degree equations in radicals for $n \geq 5$, to the seminal works by A. L. Cauchy, N. H. Abel, and finally Galois.[14]

Galois's works[15] marked a turning point, providing answers to the open questions in the solutions of equations by using new methods and algebraic notions, first of all the notion of a *group*. This notion was introduced by Galois in relation to the properties of the set of permutations of the roots of equations (the permutations constituting what he named a 'group'), together with other basic notions of group theory such as "subgroup", "normal subgroup", and "simple group".[16] By characterizing an equation in terms of its "degree of symmetry", determined by the permutation group of the roots preserving their algebraic relations (later known as the *Galois group* of the equation), Galois could transform the problem of the resolution of equations into that of studying the properties of the permutation groups involved. In this way he obtained, among other things, the necessary and sufficient conditions for solving equations by radicals.

Galois's achievements in group theory, first brought to publication by Joseph Liouville in 1846, were collected and expanded in Camille Jordan's 1870 *Traité des substitutions et des équations algébriques*. Jordan's *Treatise*, the first systematic textbook on group theory, had a decisive influence on the application of this new theory, including its application to other domains of mathematical science, such as geometry and mathematical physics.

3.2 Applications of group theory: the contributions of Klein and Lie

Projective geometry, the theory of invariants and group theory: Klein and Lie's starting point

In the same year as the appearance of Jordan's *Treatise*, Sophus Lie and Felix Klein, two young mathematicians who were to become the key figures in extending the domain of application of group theory, moved for a period from Berlin to Paris to enter into contact with the French school of mathematics. Lie and Klein had just written a joint paper investigating the properties of some curves in terms of the groups of projective transformations leaving them invariant. In fact they were drawn to Paris mostly by their interest in *projective geometry*, the science founded by J. V. Poncelet to study the properties of figures preserved under central projections. Projective geometry had become, at the time, a particularly fruitful research field for the combination of algebraic and geometrical methods based on the notion of invariance. The *theory of invariants* was itself a flourishing branch

[14]Cauchy (1789—1857) generalized Ruffini's results in 1815; Abel (1802—1829) published in 1824 a proof of the impossibility of solving the quintic equation by radicals and in 1826 the paper *Démonstrations de l'impossibilité de la résolution algébrique des équations générales qui passent le quatrième degré.*

[15]A few "m'emoires" submitted to the Académie des Sciences, three brief papers published in 1830 in the Académie's 'Bulletin', and some letters, among which is the last one written to his friend Auguste Chevalier in the night before the fatal duel.

[16]See [Yaglom, 1988, 9 f.], for details.

of mathematics, centered on the systematic study of the invariants of "algebraic forms". Using the theory of invariants, the English algebraist Arthur Cayley[17] had recently clarified the relationship between Euclidean and projective geometry, showing the former to be a special case of the latter. Before leaving for France, Klein had tried to extend Cayley's results, based on the possibility of defining a distance (a "metric") in terms of a quadratic form defined on the projective space, to the case of non-Euclidean geometries.

While in Paris Lie and Klein became acquainted not only with Jordan, but also with the expert in differential geometry Gaston Darboux, who stimulated their interest in the relations between differential geometry and projective geometry.

Klein's Erlangen Program

The question of the relations between the different contemporary geometrical systems particularly interested Klein. He aimed at obtaining a unifying foundational principle for the various branches into which geometry had apparently recently separated. In this respect, he fruitfully combined (a) the application of the theory of invariants to the study of geometrical properties, with (b) his and Lie's idea of applying algebraic group theory to treating also geometrical transformations. The new group theoretical conception of a geometrical theory which resulted was announced in his famous *Erlangen Program*, as it became known following the inaugural lecture entitled 'Comparative Considerations on Recent Geometrical Research'[18] that the 23-year-old Klein delivered when entering, in 1872, as a professor on the staff of the University of Erlangen. Guided by the idea that geometry is in the end a unity, Klein's solution to the problem posed by the existence of different geometries was to propose a general characterization of a geometrical theory by using the notion of invariance under a transformation group (i.e., the notion of symmetry). According to his characterization a geometry is defined, with respect to a given domain (the plane, the space, or a given "manifold") and a group of transformations acting on it, as the science studying the invariants under the transformations of the group. Each specific geometry is thus determined by the characterizing symmetry group (for example, planar Euclidean geometry is determined by the group of affine transformations acting on the plane), and the interrelations between geometries can be described by the relations between the corresponding groups (for example, the equivalence of two geometries amounts to the isomorphism between the corresponding groups).

With Klein's definition of a geometry, geometrical and symmetrical properties become very close: the symmetry of a figure, which is defined in a given "space"[19] the "geometrical" properties of which are preserved by the transformations of a group $\mathcal{G}$, is determined by the subgroup of $\mathcal{G}$ leaving the figure invariant. The new

[17]Cayley was one of the three members of the 'invariant trio', as the French mathematician Hermite dubbed them, the other two being James Sylvester, inventor of most of the terminology of the theory including the word 'invariant', and George Salmon.

[18]'Vergleichende Betrachtungen über neuere geometrische Vorschungen'.

[19]A set of points endowed with a structure.

group theoretical techniques prompted a transition from an inductive approach (familiar from the nineteenth century classifications of crystalline forms in terms of their visible — and striking — symmetry properties)[20] to a more abstract and deductive approach. This is the procedure formulated in Weyl's classic book on symmetry (*Symmetry*, [1952]) as follows:

> *Whenever you have to do with a structure-endowed entity* Σ *try to determine its group of automorphisms*, the group of those element-wise transformations which leave all structural relations undisturbed. [...] After that you may start to investigate symmetric configurations of elements, i.e. configurations which are invariant under a certain subgroup of the group of all automorphisms [Weyl, 1952, 144, emphasis in original].

In this way, the symmetry classifications could be extended to figures in "spaces" different from the plane and space of common experience.

Klein himself contributed to the classification of symmetry groups of figures with his works on discrete groups; in particular, he studied the transformation groups related to the symmetry properties of regular polyhedra, which proved to be useful in the solution of algebraic equations by radicals.

Lie's theory of continuous groups

After 1872, while Klein was concerned especially with *discrete* transformation groups, Lie devoted all his research work to building the *theory of continuous transformation groups*, the results of which were systematically collected in his three-volume *Theorie der Transformationsgruppen* [I: 1888, II: 1890, and III: 1893], written with the collaboration of F. Engel. Lie's interest in continuous groups arose in relation to the theory of differential equations, which he took to be 'the most important discipline in modern mathematics'. By the time he was in Paris, Lie had begun to study the theory of first-order partial differential equations, a theory of particular interest because of the central role it played in the formulation given by W. R. Hamilton and C. G. Jacobi to mechanics.[21] His project was to extend to the case of differential equations Galois's method for solving algebraic equations: that is, using the knowledge of the 'Galois group' of an equation (the symmetry group formed by the transformations taking solutions into solutions) so as to solve it or reduce it to a simpler equation. Thus Lie's guiding idea was that continuous transformation groups could, in the solution of differential equations, play a role analogous to that of the permutation groups used by Galois in the case of algebraic equations.

[20] A classic textbook in this respect is [Shubnikov and Koptsik, 1974]. See also Section 8, below.

[21] For details on classical mechanics we refer the reader to [Butterfield, this vol., ch. 1] and the references therein.

Lie had already considered continuous groups of transformations in some earlier geometrical works. In his studies with Klein on special kinds of curves (called by them 'W-curves'), he had examined transformations that were continuously related in the sense that they were all generated by repeating an *infinitesimal transformation*.[22] The relevance of infinitesimal transformations to continuous groups of transformations was to become a central point in his studies of *contact transformations*, so called because they preserved the contact or tangency of surfaces. Lie had started to investigate contact transformations in association with geometrical reciprocities implied in his "line-to-sphere mapping", a mapping between a line geometry and a sphere geometry that he had discovered while in Paris.[23] When he turned to considering first-order partial differential equations, he soon realized that they admitted contact transformations as symmetry transformations (i.e. transformations taking solutions into solutions). Thus contact transformations could form the "Galois group" of first-order partial differential equations. This motivated him to develop the invariant theory of contact transformations, which represented the first step of his general theory of continuous groups.

Lie's crucial result, allowing him to pursue his program, was the discovery that to each continuous transformation group could be assigned what is today called its *Lie algebra*. Lie showed that the infinitesimal generators of a continuous transformation group obey a linearized version of the group law, involving the commutator bracket (or Lie bracket); this linearized law then represents the structure of the algebra. In short (and in modern terminology): we describe the elements (transformations) of a continuous group (now called a *Lie group*)[24] as functions of a certain number r of continuous parameters a_l $(l = 1, 2, ...r)$. And these group elements can be written in terms of a corresponding number r of infinitesimal operators X_l, the *generators* of the group, which satisfy the "multiplication law" represented by the Lie brackets

$$[X_s, X_t] \;=\; c^q_{st} X_q \;\;,$$

so forming what is called the *Lie algebra* of the group. The coefficients c^q_{st} are constants characterizing the structure of the group and are called the *structure constants* of the group.[25]

Thanks to this sort of result, the study and classification of continuous groups could be conducted in terms of the corresponding Lie algebras. This proved to be extremely fruitful in the successive developments, not only algebraic and geometrical, but also physical. With regard first of all to the physics of Lie's time, Lie had arrived at the correspondence between continuous groups and Lie algebras by reinterpreting, in the light of his program for solving differential equations, the

[22]See on this part [Hawkins, 2000, Section 1.2]. According to Hawkins (p. 15), with the works of Lie and Klein on W-curves 'for the first time not only is a continuous group the starting point for an investigation, but also for the first time in print we have the idea that infinitesimal transformations are a characteristic and useful feature of continuous systems of transformations'.

[23]See [Hawkins, 2000, Section 1.4].

[24]For a precise definition of this and other terms in this paragraph, see [Butterfield, this vol., ch. 1, Section 3].

[25]For more details, see [Butterfield, this vol., ch. 1, Sections 3.2 and 3.4].

results obtained by Poisson and especially Jacobi about the integration of first-order partial differential equations arising in mechanics.[26] His achievements were thus of great relevance to the solution of the dynamical problems discussed by his contemporaries.

But it is in twentieth century physics, with the works of such figures as Hermann Weyl, Emmy Noether and Eugene Wigner (just to recall the central figures who first contributed to the applications of Lie's theory to modern physics), that the theory of Lie groups and Lie algebras acquired a fundamental role in the description of physical phenomena. Today, the applications of the theory that originated from Lie's works include the whole of theoretical physics, of both the large and the small: classical and quantum mechanics, relativity theories, quantum field theory, and string theory.[27]

4 WHAT ARE SYMMETRIES IN PHYSICS? DEFINITIONS AND VARIETIES

4.1 What is meant by 'symmetry' in physics

We can understand intuitively the generalization of the scientific notion of symmetry from physical or geometric objects to laws, as follows. We write down our law as a mathematical equation, and appearing in this equation will be various mathematical objects and operators. For a particular group of transformations, these objects and operators transform according to rules that may be fixed either by the mathematical nature of the object or operator concerned, or (where the mathematics does not fix the transformation rules) by our specification. If the "form" of the equation is preserved when we transform each of the objects and operators appearing in our equation by any element of the group, then we say that the group is a *symmetry group* of the equation.

More precisely, what we mean by the symmetry transformations of the laws in physics can be formulated in either of the following ways, which are equivalent in the sense that they pick out the same set of transformations:

(1) Transformations, applied to the independent and dependent variables of the theory in question, that leave the form of the laws unchanged.

(2) Transformations that map solutions into solutions.

Symmetry transformations may be viewed either *actively* or *passively*. From the passive point of view we re-describe *the same physical evolution* in two different coordinate systems.[28] That is, we transform the independent and dependent variables, as in (1). If the description in the original set of coordinates is a solution

[26] For details see [Hawkins, 2000, Section 2.5].

[27] In this volume, see especially t'Hooft (ch. 7), Dickson (ch. 4), and Belot (ch. 2).

[28] By 'coordinates' here we are referring to generalized coordinates; in general, one coordinate for each degree of freedom of the system.

of given equations, then the new description in the new set of coordinates is a solution *of the same equations*. (If the transformation is *not* a symmetry transformation, then the new description in the new coordinates will not, in general, be a solution of the same equations, but rather of *different* equations.) The mapping of one solution into another solution of the same equations, by means of a symmetry transformation, leads to the *active* interpretation of such transformations. On this interpretation, the two solutions are viewed as *different physical evolutions* described in *the same coordinate system*. Thus, formulation (2) lends itself naturally to an active interpretation.

The 'form of the law' in (1) means the functional form of the law, expressed in terms of the independent and dependent variables. A transformation of those variables will, in general, lead to an expression whose functional form differs from that of the original expression (x goes to x^2, for example). At this point it will be helpful to say a few words about "invariance" and "covariance". Let the reader beware that there is no unanimity over how these terms are used in discussing the laws of physics, especially in the philosophy of physics literature. Often, the term 'invariant' is reserved for objects, and 'covariant' is used for equations or laws. However, this is a product of a more fundamental distinction, which when understood correctly allows for the application of the notion of invariance to laws as well.

We think that the discussion of Ohanian and Ruffini [1994, Section 7.1] is very useful, and that it nicely distils much of the best of what can be found in the literature, both in physics and in philosophy of physics. The upshot is as follows. We may say that an equation is *covariant* under a given transformation when its form is left unchanged by that transformation. This is the notion at work in Definition 1. In a way, it is rather weak: given an equation that is not covariant under a given transformation, we can always re-write it so that it becomes covariant. On the other hand, this re-writing may involve the introduction of new functions of the variables, and it is the physical interpretation of these new quantities that allows covariance to gain physical significance. We will have more to say about this for the specific case of *general covariance* and Einstein's General Theory of Relativity in Section 6.3 below.

Invariance of an equation, as characterized by Ohanian and Ruffini, is a stronger requirement than covariance. Not only should the form of the equation remain the same, but so too should the values of any non-dynamical quantities, including "constants" such as the speed of light. By "non-dynamical quantities" we mean all those objects which appear in the equations yet which do not themselves satisfy equations of motion. We here enter the muddy waters of how to distinguish between "absolute" and "dynamical" objects, as discussed by Anderson [1967].[29]

In both cases (covariance and invariance), the associated transformations — when actively construed — take solutions into solutions. When using formulation

[29]See also Section 6.3, below. One difficulty in tackling the literature on this issue is the variety of uses and meanings attaching to the common terminology of covariance, principle of covariance, invariance, absolute and dynamical objects, and so forth.

(2), it is important to be clear about what is meant by a solution. This does *not* mean a solution-at-an-instant, i.e. an instantaneous state of a system; rather, it means an entire history, i.e. possible time-evolution, of the system in question.[30]

4.2 Varieties of symmetry

Symmetries in physics come in a number of different varieties, distinguished by such terms as 'global' and 'local'; 'internal' and 'external'; 'continuous' and 'discrete'. In this Section we briefly review this terminology and the associated distinctions.

The most familiar are the global spacetime symmetries, such as the Galilean invariance of Newtonian mechanics, and the Lorentz invariance of the Special Theory of Relativity. Global spacetime symmetries are intended to be valid for all the laws of nature, for all the processes that unfold in the spacetime. Symmetries with this universal character were labelled 'geometric' by Wigner (see [1967, especially p. 17]).

This universal character is not shared by some of the symmetries introduced into physics during the twentieth century. Most of these were of an entirely new kind, with no roots in the history of science, and in some cases expressly introduced to describe specific forms of interactions — whence the name 'dynamical symmetries' due to Wigner [1967, see especially pp. 15, 17—18, 22—27, 33]).

The various symmetries of modern physics can also be classified according to a second distinction: that between global and local symmetries. The terms 'global' and 'local' are used in physics, and in philosophy of physics, with a variety of meanings. The distinction intended here is between symmetries that depend on constant parameters (global symmetries) and symmetries that depend on arbitrary smooth functions of space and time (local symmetries). While Lorentz invariance is an example of a global symmetry, the gauge symmetry of classical electromagnetism (an internal symmetry)[31] and the diffeomorphism invariance in General Relativity (a spacetime symmetry) are examples of local symmetries, since they are parameterized by arbitrary functions of space and time.[32] Recalling Wigner's distinction, Lorentz invariance is a geometric symmetry, applying to all interactions, whereas the gauge symmetry of electromagnetism concerns the electromagnetic interaction specifically and is therefore a dynamical symmetry.

The gauge symmetry of classical electromagnetism is an internal symmetry because the transformations of the vector potential occur in the internal space of the field system, rather than in spacetime. The gauge symmetry of classical electromagnetism can seem to be no more than a mathematical curiosity, specific to this theory; but with the advent of quantum theory the use of internal degrees

[30] The distinction is important in, for example, our discussion of Curie's principle, Section 2.2 above.

[31] For more on gauge and internal symmetries, see the following paragraph.

[32] We discuss the local symmetry of General Relativity further in Section 6.1 below. See also [Belot, this vol., ch. 2].

of freedom, and the related internal symmetries, became fundamental.[33]

The translations, rotations and boosts of the inhomogeneous Lorentz group are all examples of *continuous symmetries*, for which any finite symmetry transformation can be built up of infinitesimal symmetry transformations. In contrast with the continuous symmetries we have the *discrete symmetries* of charge conjugation, parity, and time reversal (CPT), along with permutation invariance. Thus, Newtonian mechanics and classical electrodynamics are invariant under parity (left-right inversion) and under time reversal (roughly: the laws hold for a sequence of states evolved in the backwards time direction just as they hold for the states ordered in the forwards direction). Classical electrodynamics is also invariant under charge conjugation, so long as we correctly implement the associated transformations of the electric and magnetic fields. Finally, there is a sense in which classical statistical mechanics is permutation invariant: the particles postulated are identical to one another, and their permutation takes a solution into a solution. However, the power and significance of the discrete symmetries achieves its full force only in quantum theory.

In Section 8 below, we discuss some of the interpretative issues associated with these different varieties of symmetry in classical physics.

5 SOME APPLICATIONS OF SYMMETRIES IN CLASSICAL PHYSICS

5.1 Transformation theory in classical mechanics

As we have seen, Lie's interest in continuous groups arose in relation to his studies of the theory of first-order partial differential equations, which played a central role in the formulation given by Hamilton and Jacobi to mechanics. The *transformation theory of mechanics* based on this formulation is indeed one of the first examples of a systematic exploitation in physics of the invariance properties of dynamical equations. These symmetries are exploited according to the following strategy: the integration of the equations of motion is simplified by transforming — by means of symmetry transformations — the original dynamical system into another system with fewer degrees of freedom.

Historically, the road to the possibility of applying the above 'transformation strategy' to solving dynamical problems was opened by the works of J. L. Lagrange and L. Euler. The Euler-Lagrange analytical formulation of mechanics, grounded in the seminal *Mécanique Analytique* [1788] of Lagrange, expressed the laws of motion in a form which was covariant (cf. Section 4.1) under all coordinate transformations. This meant one could more easily choose coordinates to suit the dynamical problem concerned. In particular, one hoped to find a coor-

[33]For interpretative issues associated with gauge symmetry in classical electromagnetism, see Belot [1998]. Gauge symmetries came to prominence with the development of quantum theory. The term 'gauge symmetry' itself stems from Weyl's 1918 theory of gravitation and electromagnetism. For discussions of all these aspects of gauge symmetry, see [Brading and Castellani, 2003].

dinate system containing "cyclic" (a.k.a. "ignorable") coordinates. The presence of ignorable coordinates amounts to a partial integration of the equations: if all the coordinates are ignorable, the problem is completely and trivially solved. The method was thus to try to find (by applying coordinate transformations leaving the dynamics unchanged) more and more ignorable variables, thus transforming the problem of integrating the equations of motion into a problem of finding suitable coordinate transformations.

The successive developments in the analytical approach to mechanics, from Hamilton's "canonical equations of motion" to the general transformation theory of these equations (the theory of canonical transformations) obtained by Jacobi, presented many advantages of the "transformation strategy" point of view. For further details we refer the reader to Butterfield's chapter of this volume, along with classic references such as Lanczos [1949; 1962] and Whittaker [1904; 1989]. Butterfield [this vol., ch. 1], by expounding the theory of symplectic reduction in classical mechanics, thoroughly illustrates the strategy of simplifying a mechanical problem by exploiting a symmetry. This strategy is also the main subject of Butterfield [2006], focussing on how symmetries yield conserved quantities according to Noether's first theorem (see Section 7, below), and thereby reduce the number of variables that need to be considered in solving a problem.

We end these brief remarks on symmetry and transformation theory in classical mechanics by emphasizing two points.

First, we note that a problem-solving strategy according to which a dynamical problem (equation) is transformed into another equivalent problem (equation) by means of a symmetry might be seen as an example of the application of Curie's principle in its modern version (see here Section 2.2): by transforming an equation into another equivalent equation using a specific symmetry we may arrive at an equation which we can solve; the solution of the new equation is related to the unknown solution of the old equation by the specific symmetry; that is, we thereby arrive at an equivalent solution.

Second, we emphasize that in all these developments the invariance properties of the dynamical equations, though undoubtedly important, were considered exclusively in an instrumental way. That is, canonical transformations were studied only for the purpose of solving the dynamical problem at hand. The equations were given, and their invariance properties were investigated to help find their solutions. The formulation of Einstein's Special Theory of Relativity at the beginning of the twentieth century brings an inversion of this way of thinking about the relationship between symmetries and physical laws, as we shall see in the following section.

5.2 Symmetry principles as guides to theory construction

The principle of relativity, as expressed by Einstein in his 1905 paper announcing the Special Theory of Relativity, asserts that

> The laws by which the states of physical systems undergo changes are

> independent of whether these changes are referred to one or the other of two coordinate systems moving relatively to each other in uniform motion.[34]

It further turns out that these coordinate systems are to be *inertial* coordinate systems, related to one another by the Lorentz transformations comprising the inhomogeneous Lorentz group.

The principle of relativity thus stated meets the conditions listed above in Section 1 for a *symmetry* principle:

- The Lorentz transformations, applied to the independent and dependent variables of the theory, leave the form of the laws as stated in one inertial system unchanged on transformation to another inertial coordinate system.

- The Lorentz transformations map a solution, given relative to an inertial coordinate system, into another solution.

This principle was explicitly used by Einstein as a guide to theory construction: it is a principle that must be satisfied whatever the final details of the theory.[35] Indeed, using just the principle of relativity and the light postulate, Einstein derives various results, including the Lorentz transformations. As noted above, this represents a reversal in the priority that, since the time of Newton, had been given to the relativity principle versus the dynamical laws. Huygens used the relativity principle as a basic postulate from which to derive dynamical results, but in Newton the relativity principle, initially presented in his manuscripts as an independent postulate, is relegated in the *Principia* to a corollary.[36] From then until Einstein, the relativity of inertial motion is seen as a consequence of the particular laws under consideration, and something that could turn out to be false once the details of the laws of some particular interaction are known. Similarly for classical physics in general, symmetries — such as spatial translations and rotations — were viewed as properties of the laws that hold as a consequence of those particular laws. With Einstein that changed: symmetries could be postulated prior to details of the laws being known, and used to place restrictions on what laws might be postulated. Thus, symmetries acquired a new status, being postulated independently of the details of the laws, and as a result having strong heuristic power. As Wigner wrote, Einstein's papers on special relativity 'mark the reversal of a trend': after Einstein's works, 'it is now natural to try to derive the laws of nature and to test their validity by means of the laws of invariance, rather than to derive the laws of invariance from what we believe to be the laws of nature' [Wigner, 1967, 5].

---

[34] Miller's [1981] translation, p. 395.

[35] For discussion of the principle/constructive theory distinction in Einstein, see [Brown, 2006, ch. 6] and [Howard, 2007].

[36] In fact, it does not follow from Newton's three laws of motion — we must further assume the velocity independence of mass and force. See [Barbour, 1989, Section 1.2].

The methodology that had served Einstein well with the Special Theory of Relativity (STR) also had a role in his development of the General Theory (GTR), for which he used various different principles as restrictions on the possible form that the eventual theory might take.[37] One of these was, so Einstein maintained, an extension of the principle of relativity found in STR to include coordinate systems that are in accelerated motion relative to one another, implemented by means of the requirement that the equations of his new theory be generally covariant. Einstein was seeking a "Machian" solution to the challenge of Newton's bucket, which he took to require that there be no preferred reference frames. Thus, in his 1916 review article Einstein wrote that '*The laws of physics must be of such a nature that they apply to systems of reference in any kind of motion.* Along this road we arrive at an extension of the postulate of relativity' (emphasis in original).

The questions of whether or not the principle of general covariance (a) makes any arbitrary smooth coordinate transformation into a symmetry transformation, and (b) is a generalization of the principle of relativity, have been much discussed. The answer to (b) is a definitive 'no', but there is less consensus at present about the answer to (a).[38] In the following section we take up discussion of (a). Here we close with a few brief remarks concerning (b).

Even if general covariance in GTR is a symmetry principle, it is not an extension of the relativity principle. That is to say, general covariance says nothing about the observational equivalence of distinct reference frames.[39] As already noted, the thought that general covariance might provide such a principle was, for Einstein, connected with his attempts to provide a "Machian" resolution to the challenge of Newton's bucket, and with his principle of equivalence. However, the principle of equivalence *does not* imply the observational equivalence of reference frames in arbitrary states of motion (Einstein never thought that it did), and Einstein eventually realized that GTR does not vindicate a solution to Newton's bucket that depends only on the relative motion of *matter*.[40]

Whatever the subtleties of whether, and to what extent, general covariance is a symmetry principle, it is clear that it had enormous heuristic power, not just for Einstein in his development of GTR, but also beyond. Think for example Hilbert's work on the axiomatization of physics (see [Corry, 2004], and references therein), and Weyl's attempts to construct a unified field theory (see [O'Raifeartaigh, 1997], for an English translation of Weyl's 1918 paper 'Gravitation and Electricity', and see also [Weyl, 1922]). In all these cases, general covariance provided a powerful tool for theory construction. In the following Section we discuss further the

[37]Primarily the following: the principle of relativity, later (in 1918) distinguished from what Einstein referred to as 'Mach's principle'; the principle of equivalence; and the principle of conservation of energy–momentum.

[38]See for example [Torretti, 1983, 152–4]; Norton [1993], who also discusses the relationship with the principle of equivalence; Anderson [1967].

[39]For further discussion see, for example, [Norton, 1993] and [Torretti, 1983, Section 5.5].

[40]For a clear and concise discussion, see [Janssen, 2005].

significance and interpretation of general covariance in GTR.[41]

6 GENERAL COVARIANCE IN GENERAL RELATIVITY

In the preceding Section we noted the role of the principle of general covariance as a guide to theory construction. In this Section we turn our attention to a number of further issues relating to general covariance in GTR that have received attention in the philosophical literature. We begin with the issue, raised in the preceding section, concerning the status of arbitrary smooth coordinate transformations as symmetry transformations. We then discuss various characteristics associated with general covariance, including those pointed to by Einstein's so-called 'hole argument', before turning to the issue of whether or not general covariance has physical content.[42] We postpone discussion of Noether's theorems to Section 7, below.

6.1 General covariance and arbitrary coordinate transformations as symmetry transformations

Does the principle of general covariance make any arbitrary smooth coordinate transformation into a symmetry transformation? One way to approach this question is to consider active rather than passive transformations (see Section 4.1, above), and to compare the situation in GTR with that in STR.

In STR, a Lorentz transformation — actively construed — picks up the matter fields and redistributes them with respect to the spacetime structure encoded in the metric. The principle of relativity holds for such transformations because the evolution of the matter fields in the two cases (related by the Lorentz transformation) are observationally indistinguishable: no observations, in practice or in principle, could distinguish between the two scenarios. In GTR, active general covariance is implemented by active diffeomorphisms on the spacetime manifold (see [Rovelli, this vol., ch. 12, Section 4.1]). These involve transformations of not just the matter fields, but also the metric field, in which both are redistributed with respect to the spacetime manifold. Once again, the "two cases" are observationally "indistinguishable", but this time the reason generally given is that the "two cases" are in fact just one case.[43]

Why should we accept that there are two genuinely distinct cases when considering the Lorentz transformations in STR, and only one case for the diffeomorphisms of GTR? One approach would be to claim that a crucial difference between the two is that a Lorentz transformation can be implemented on an effectively isolated *sub-system* of the matter fields, producing an observably distinct scenario in

[41]For detailed presentation of the Special and General Theories of Relativity, see [Malament, this vol., ch. 3]. See also [Rovelli, this vol., ch 12, Section 4].

[42]See also [Belot, this vol., ch. 2].

[43]See also Section 6.2, penultimate paragraph.

which, nevertheless, the evolution of the sub-system in question is indistinguishable assuming no reference is made to matter fields outside that subsystem. For example, in Galileo's famous ship experiment we consider two observably distinct scenarios — one in which the ship is at rest with respect to the shore, and one in which it moves uniformly with respect to the shore — and we notice that the behaviour of physical systems within the cabin of the ship does not distinguish between the two scenarios.[44] No analogue of the Galilean ship experiment can be generated for the general covariance of GTR.[45]

The importance of symmetry transformations being implementable to produce observationally distinct scenarios has been emphasized by Kosso [2000]. On this view, the observational significance of symmetry transformations rests on a combination of *two* observations being possible in principle. First, it must be possible to confirm empirically the implementation of the transformation — hence the importance of being able to generate an observationally distinguishable scenario through the transformation of a subsystem. Second, we must be able to observe that the subsequent internal evolution of the subsystem is unaffected. That we cannot meet the first of these requirements for arbitrary smooth coordinate transformations in GTR marks a difference between these and the Lorentz transformations.[46]

On this approach, while the field equations of GTR take the same form for any choice of coordinate system, this is not sufficient for arbitrary coordinate transformations to be symmetries. In addition, the actively construed transformations must have a physical interpretation — we must be transforming one thing with respect to something else. When we perform a diffeomorphism, we get back the same solution, not a new solution, for we are not re-arranging the matter fields with respect to the metric.

We stress that this is only one way to approach the issue of whether general covariance should be understood as a symmetry principle in GTR. A contrasting position may be found in [Anderson, 1967, Section 10-3], who argues that we must understand Einstein as viewing general covariance as a symmetry requirement, and attempts to spell out the conditions under which it can function as such.

[44] This implementation can be only approximate, relying on the degree to which the subsystem in question can be isolated from the "external" matter fields.

[45] One suggestion might be that we perform a transformation T which is the identity outside some region R, and which differs from the identity within that region. This will not achieve the desired result. The two scenarios must have observationally distinct consequences, at least in principle. In the case of Galileo's ship, if we allow the subsystem to interact with other matter once again, we will see that in one case the ship crashes into rocks (for example), while in the other it suffers no such collision. Thus, we have observational distinguishability in principle. The transformation T does not produce a scenario which any future events could enable us to distinguish from the original.

[46] Indeed, this result applies generally to local versus global symmetries. See also [Brading and Brown, 2004].

6.2 Characteristics of generally covariant theories

Any generally covariant theory will possess certain characteristics that are philosophically noteworthy. First, there will be a *prima facie* problem with causality and determinism within the theory, and second, there will be constraints on the specification of the initial data. Einstein recognized aspects of the first characteristic while he was searching for his theory of gravitation, maintaining from 1913 through until the fall of 1915 that his so-called 'hole argument' provided grounds for concluding that no generally covariant theory could be physically acceptable.

In the 'hole argument', Einstein considers a region of spacetime in which there are no matter fields (the "hole"), and then shows that in a generally covariant theory no amount of data about the values of the matter and gravitational fields outside the hole is sufficient to uniquely determine the values of the gravitational field inside the hole. From this, Einstein concluded that no generally covariant theory could be physically acceptable.[47]

The context to bear in mind here is that Einstein was searching for a theory in which the matter fields plus the field equations would uniquely determine the metric.[48] In the summer of 1915 Einstein lectured on relativity theory in Göttingen where his audience included David Hilbert. If we assume that Einstein's presentation included a version of his 'hole argument', then we can reasonably infer that Hilbert was quick to reinterpret the issue that the 'hole argument' points to, and to present the problem raised for generally covariant theories in terms of whether such theories permit well-posed Cauchy problems.[49]

In the years immediately following the advent of GTR, Hilbert played a central role in spelling out the problems of causality and determinism faced by any generally covariant theory. He pointed out that in any such theory, including GTR, there will be four fewer field equations than there are variables, leading to a mathematical underdetermination in the theory. As Hilbert stressed, the Cauchy problem is not well-posed: given a specification of initial data, the field equations do not determine a unique evolution of the variables.

We can see the connection between the underdetermination problem and general covariance as follows. For the Cauchy problem to be well-posed, we must be able to express the second time derivatives of the metric in terms of the initial data (plus the further spatial derivatives that can be calculated from the initial data). However, if we re-express the 10 (source-free) Einstein field equations $G_{\mu\nu} = 0$ so as to explicitly display all the terms containing the second time derivative of the metric, we see that we have ten equations for six unknowns $g_{ij,00}$, the remaining four second time derivatives $g_{\mu 0,00}$ failing to appear in the equations.[50] This

---

[47]For presentation and discussion of the 'hole argument', see Norton [1984, 286–291] and [1993, Sections 1-3], Stachel [1993], and Ryckman [2005, Section 2.2.2]. See also [Rovelli, this vol., ch. 12, Section 4.1.1].

[48]For more on Einstein's (mis)appropriation of Mach's principle, see [Barbour, 2005].

[49]Brading and Ryckman [2007]; see also [Brown and Brading, 2002, especially Section IV].

[50]See [Adler, Bazin and Schiffer, 1975, ch. 8] for details of the over- and under-determination issues.

is a direct consequence of general covariance: we can always make a coordinate transformation in the neighborhood of the initial data surface such that the metric components and their first derivatives are unchanged, while the second time derivatives $g_{\mu 0,00}$ vanish on that surface. Thus the field equations, which must be valid in all coordinate systems, cannot possibly contain information on the second time derivatives. The initial data do not determine the metric uniquely: there are four arbitrary functions $g_{\mu 0,00}$ that we are free to choose.

Today, it is customary to assert from the outset that solutions of Einstein's field equations differing only in the choice of these four arbitrary functions are physically equivalent.[51] But here we should note that this "gauge freedom" interpretation of general covariance leads to problems of its own.[52] For example, within this framework the observables of the theory must be "gauge invariant" quantities, but such quantities have (to date) turned out to be far removed from anything "observable" in the operational sense. The gauge freedom interpretation of general covariance is sometimes accompanied by the view that this freedom — and therefore general covariance itself — lacks physical content. We turn to consider this issue in Section 6.3, below.

In our explanation of the underdetermination problem, above, we noted that the Einstein equations provide ten equations for the six unknowns $g_{ij,00}$. The other face of the underdetermination problem is therefore an *overdetermination* problem with respect to the $g_{ij,00}$, and what this means is that there will be constraints on the specification of the data on the initial hypersurface. This is the second characteristic of all generally covariant theories that we mentioned in our opening remarks of the current subsection. Indeed, the presence of constraint equations is a feature shared with other theories with a local symmetry structure, such as electromagnetism. Philosophically, the significance lies in the relationship between the theory and the initial data. In the seventeenth century Descartes wrote a story of a world created in a state of disorder from which, by the ordinary operation of the laws of nature, a world seemingly similar to our own emerged.[53] This image of the world emerging from an initial chaos has a long history, of course, but the emergence of order by means of the operation of the laws of nature offered a novel twist to the tale. It involves the separation of initial conditions, which could be anything, from the subsequent law-governed evolution of the cosmos. In modern terms, this is a theory without constraints: the theory determines *which* properties of a system must be specified in order to give adequate initial data, but we are then free to assign whatever *values* we please to these properties; the equations of the theory are used to evolve that data forwards in time. A theory with constraints, by contrast, contains two types of equations: constraint equations that must be satisfied by the initial data, as well as evolution equations.

[51] Recall the discussion of Section 6.1, above.

[52] See [Belot, this vol., ch. 2].

[53] Written around 1633, *Le Monde* was not published in Descartes's lifetime. For an English translation see Descartes [1998]. The "order out of disorder" story is in the Treatise on Light, chs. 6 and 7. Whether the ordinary operation of the laws of nature was sufficient to bring order out of chaos became a much-disputed issue.

In GTR, four of the ten field equations connect the curvature of the initial data hypersurface with the distribution of mass–energy on that hypersurface, and the remaining six field equations are evolution equations. To sum up, in a theory with constraints, the initial "disorder" cannot be so disordered after all, but must itself satisfy constraints set down by the laws of the theory.

6.3 Does general covariance have any physical significance?

As we saw in Section 5, Einstein treated general covariance as a symmetry principle guiding the search that produced his General Theory of Relativity. There is no doubt that general covariance proved a useful heuristic for Einstein, but there remains an ongoing dispute over whether general covariance in fact has any physical significance. The issue was forcefully raised by Kretschmann already in 1917. The thrust of the argument, which continues to reverberate today, is that any theory can be given a generally covariant formulation given sufficient mathematical ingenuity, and therefore the principle of general covariance places no restrictions on the physical content of a theory. Indeed, Norton [2003] begins his discussion of the issue by claiming that this negative view of general covariance has become mainstream, before going on to give an alternative viewpoint (see below).

It seems clear to us that the characteristic features of generally covariant theories discussed above may, in some theories at least (including GTR), be far from trivial, and that the mainstream view — which would indeed render these issues trivial — should be opposed. Those wishing to oppose the mainstream view adopt a two-step general strategy: first, show under what conditions general covariance places a restriction on the physical content of a theory; and second, demonstrate what those implications for physical content consist in. Thus, the general mathematical point that any theory can be put into generally covariant form is conceded, but the implication that general covariance is therefore necessarily physically vacuous is resisted by attention to the manner in which general covariance is implemented in a given theory or class of theories.

For example, Anderson [1967], Ohanian and Ruffini [1994], Norton [2003], and Earman [2006] each attempt to explain under what conditions the purely mathematical feature of general covariance comes to have physical bite.[54] Anderson distinguishes between the symmetries of a theory (which have physical significance) and the covariance group of the equations (which need not). Anderson is the classic reference for the distinction between "absolute" and "dynamical" objects,[55]

[54] See also [Norton, 1993, especially Section 5], and [Rovelli, this vol., ch. 12, Section 4.1.3].

[55] It has proved difficult to make the distinction between absolute and dynamical objects precise, but the intuitive idea is clear enough. Dynamical objects satisfy field equations and interact with other objects, whereas absolute objects are not affected by the dynamical behaviour of other fields appearing in the theory. For a careful and detailed treatment of Anderson's approach, and the counter-examples that have been raised, see [Pitts, 2006]. The conclusion of this paper is that Anderson's intuition can be made sufficiently precise to cope with all counter-examples that have appeared in the literature to date (including one due to Pitts himself), but that there is another example, due to Geroch, that Pitts has been unable to resolve. The debate goes on!

and in this terminology the covariance group of the equations of a theory becomes a symmetry group if and only if the theory contains no absolute objects. Ohanian and Ruffini [1994] appeal to the distinction they make between *invariance* and *covariance* of the equations of a theory.[56] Covariance, they agree, is a mathematical feature (perhaps simply an artefact of the particular formulation of the theory at hand); but we require not only the covariance of the equations, but also that for any objects (with one or more components) appearing in the theory that are nevertheless independent of the state of matter (such as the speed of light, Planck's constant, etc.), their value should be unchanged by the general coordinate transformations. Norton [2003] emphasizes the role of physical considerations in fixing the content of a theory such that this restricts the formal games that we can play. Earman [2006] begins by taking pains to emphasize the distinction between the 'mere co-ordinate freedom' (associated with arbitrary coordinate transformations, passively construed) and 'the substantive demand that diffeomorphism invariance is a gauge symmetry of the theory at issue'. That is to say, he reminds us that the issue at stake is not our ability to re-write a theory in generally covariant form (it is conceded that this is something we can always do, given sufficient mathematical ingenuity), but the relationship between the physical situations that are related by diffeomorphisms, i.e. by (active) point transformations (see Section 6.1, above). 'Substantive general covariance' holds when diffeomorphically related models of the theory represent different descriptions of the same physical situation. The claim is that GTR satisfies substantive general covariance whereas generally covariant formulations of such theories as STR need not, and the goal is to show that this requirement provides demarcation between theories in which general covariance represents a physically significant property of the theory, and those in which it does not.[57]

Thus, Anderson, Ohanian and Ruffini, Norton, and Earman each seek to add bite to the "merely mathematical" requirement of general covariance by placing conditions on the manner in which it is implemented in the theory. Once these requirements are added, various consequences follow for the content of the theory, such as that the metric be a dynamical object. In each case, the aim is to elevate general covariance as implemented in GTR to a symmetry principle.[58]

Considerations of the significance of general covariance in theories of gravitation led to the formulation of three theorems important for the general interpretation of symmetries in physics. These theorems are due to Emmy Noether and Felix Klein, and will be discussed in the following section.

---

[56]See Section 4.1.

[57]One important tool for distinguishing genuine 'gauge theories' from those in which the local symmetry in question is merely formal is Noether's second theorem; see Section 7, below.

[58]Brown and Brading [2002] attempt to analyze in more detail, by means of Noether's theorems (see Section 7, below), what additional conditions must be added to general covariance in order to arrive at specific aspects of the content of GTR.

7 NOETHER'S THEOREMS

Any discussion of the significance of symmetries in physics would be incomplete without mention of Noether's theorems. These theorems relate symmetry properties of theories to other important properties, such as conservation laws.

Within physics, the term 'Noether's theorem' is most frequently associated with a connection between global continuous symmetries and conserved quantities. Familiar examples from classical mechanics include the connections between: spatial translations and conservation of linear momentum; spatial rotations and conservation of angular momentum; and time translations and conservation of energy. In fact, this theorem is the first of two theorems presented in her 1918 paper 'Invariante Variationsprobleme'.[59]

Before stating the two theorems, we begin with the following cautionary remark. The connection between *variational symmetries* (connected to the invariance of the action, and in terms of which Noether's theorems are formulated) and *dynamical symmetries* (concerning the dynamical laws, which is the topic of our discussion here) is subtle (see [Olver, 1993, ch. 4]). Noether herself never addressed the connection, and never used the word 'symmetry' in her paper. She discusses integrals mathematically analogous to (but generalizations of) the action integrals of Lagrangian physics, and uses variational techniques and group theory to elicit a pair-wise correspondence between variational symmetries of the integral and a set of identities.

Noether then proves two theorems, the first for the case where the variational symmetry group depends on constant parameters, and the second for the case where the variational symmetry group depends on arbitrary functions of the variables.[60] In the following statement of her theorems we use the term 'Noether symmetry' to refer to a symmetry of the field equations for which the change in the action arising from the infinitesimal symmetry transformation is at most a surface term. Using the terminology of Section 4.2, the first type of symmetry then corresponds to a *global* dynamical symmetry, and the second to a *local* one. We state the theorems in a form appropriate to Lagrangian field theory; Noether's own statement of the theorems involves no such specialization. For discussion of the first theorem in the context of finite-dimensional classical mechanics see [Butterfield, this vol., ch. 2, Section 2.1.3]. We state the theorems so that we can refer back to them to characterize the conceptual content, but for discussion of the mathematical detail of their derivation and content we refer the reader elsewhere — see especially [Olver, 1993] and [Barbashov and Nesterenko, 1983].

We can state Noether's two theorems, for a Lagrangian density L depending on the fields $\phi_i(x)$ and their first derivatives, as follows.

[59] For an English translation see [Noether, 1971].
[60] See [Brading and Brown, 2007].

Noether's first theorem

If a continuous group of transformations depending smoothly on ρ constant parameters ω_k $(k = 1, 2, ..., \rho)$ is a Noether symmetry group of the Euler-Lagrange equations associated with a Lagrangian $L(\phi_i, \partial_\mu \phi_i, x^\mu)$, then the following ρ relations are satisfied, one for every parameter on which the symmetry group depends:[61]

$$\sum_i E_i^L \xi_i^k = \partial_\mu j_k^\mu. \tag{1}$$

On the left-hand side we have a linear combination of Euler expressions,

$$E_m^L \equiv \frac{\partial L}{\partial \phi_m} - \partial_\mu \left(\frac{\partial L}{\partial \phi_{m,\mu}} \right) \tag{2}$$

where

$$E_m^L = 0 \tag{3}$$

are the Euler-Lagrange equations for the field ϕ_m. (The ξ_i^m depend on the particular symmetry transformations and fields under consideration, and the details are not important for our current purposes.)

On the right-hand side we have the divergence of a current, j_k^μ. When the left-hand side vanishes, the divergence of the current is equal to zero, and this expression can be converted into a conserved quantity subject to certain conditions. Thus, Noether's first theorem gives us a connection between global symmetries and conserved quantities.[62]

Noether's second theorem

If a continuous group of transformations depending smoothly on ρ arbitrary functions of time and space $p_k(x)$ $(k = 1, 2, ..., \rho)$ and their first derivatives is a Noether symmetry group of the Euler-Lagrange equations associated with a Lagrangian $L(\phi_i, \partial_\mu \phi_i, x^\mu)$, then the following ρ relations are satisfied, one for every function on which the symmetry group depends:

$$\sum_i E_i^L a_{ki} = \sum_i \partial_\nu (b_{ki}^\nu E_i^L). \tag{4}$$

The a_{ki} and b_{ki}^ν depend on the particular transformations of the fields in question, and while again the details need not concern us here, we note for use below that while the a_{ki} arise even when the symmetry transformation is a global transformation, the b_{ki}^ν occur only when it is *local*.[63] What we have here, essentially,

[61]Note that we are using the Einstein summation convention to sum over repeated greek indices.

[62]This theorem is widely discussed. See especially [Barbashov and Nesterenko, 1983]; [Doughty, 1990]. We refer the reader to [Butterfield, this vol., ch. 2] and [Butterfield, 2006] for further discussion of Noether's first theorem in the context of finite-dimensional classical mechanics.

[63]Once again, the reader is referred to [Brading and Brown, 2007] for further details.

is a dependency between the Euler expressions and their first derivatives. This dependency holds as a consequence of the *local* symmetry used in deriving the theorem. In the case when all the fields are dynamical (i.e. satisfy Euler-Lagrange equations) it follows that not all the field equations are independent of one another. This *formal underdetermination* is characteristic of theories with a local symmetry structure.[64]

As Hilbert recognized in the context of generally covariant theories of gravitation, the underdetermination is independent of the specific form of the Lagrangian.[65] In the case of General Relativity, once we specify the Lagrangian and substitute it into (4), we arrive at the (contracted) Bianchi identities.

For Noether herself, the impetus for the paper arose from the discussions over the status of energy conservation in generally covariant theories between Hilbert, Klein and Einstein, during which Hilbert commented that energy conservation for the matter fields no longer has the same status in generally covariant theories as it had in previous (non-generally covariant) theories, because it follows independently of the field equations for the matter fields. Noether's two theorems can be used to support this conjecture (see [Brading, 2005]). The discussion over the status of energy conservation in General Relativity continues, the root of the issue being that energy-momentum cannot, in general, be defined locally.[66]

Today, the significance of Noether's results lie in their generality. Many of the specific connections between global spacetime symmetries and their associated conserved quantities were known before Noether's 1918 paper, and both Einstein and Hilbert anticipated some aspects of the second theorem in their investigations of energy conservation during and after the development of GTR.[67] However, her systematic treatment allows us to understand that these relations do not rely on the detailed dynamics of a particular theory, but in fact follow from the structure of Lagrangian theories and significantly weaker stipulations than the full dynamics of the theory. For example, general covariance leads to energy conservation in GTR given satisfaction of the gravitational field equations, but *independently*

---

[64]Whether the dependencies expressed by the second theorem are trivial or not depends on the status of the fields with respect to which the local symmetry holds. It is in this way that Noether's second theorem can be used as a tool in the attempt to demarcate 'true gauge theories' from theories where the local symmetry is a 'mere mathematical artefact' (see Section 6.3 above, and [Earman, 2006]). For a 'true gauge theory' the dependencies have significant physical implications.

[65][Hilbert, 1915].

[66]The energy-momentum conservation law in General Relativity is formulated in terms of the vanishing of the *covariant* divergence of the energy-momentum tensor associated with the matter fields. Alternatively, we can express this in terms of the vanishing of the coordinate divergence of the energy-momentum of the matter fields plus that of the gravitational field. The latter term falling under the divergence operator is not uniquely defined and, pertinent the issue of non-localizability, may vanish in some coordinate systems and not in others. We can understand this coordinate dependence by reflecting on the equivalence principle, according to which partitioning the inertial-gravitational field to obtain a division between inertial and gravitational forces is itself a coordinate-dependent issue. For further discussion see, for example, [Misner, Thorne and Wheeler, 1970, 467–8], and [Wald, 1984, 70]. See also [Malament, this vol., ch. 3].

[67]On Einstein, see [Janssen, 2005, 75–82]; and see [Sauer, 1999] on Hilbert.

of the detailed form of those equations, and *independently* of the field equations for the matter fields (indeed, independently of whether the matter fields satisfy Euler-Lagrange equations at all).[68] Noether's theorems are a powerful tool for investigating the structure of theories — which assumptions are required to generate which aspects of the theory, and so forth.[69]

It is worthwhile mentioning a third theorem, connected with Noether's two theorems and derived in the same context (i.e. the study of generally covariant theories of gravitation and conservation of energy) by Felix Klein [1918]. We call it the 'Boundary theorem' for reasons associated with its method of derivation.[70] As with Noether's second theorem, the Boundary theorem concerns local symmetries, and results in a series of identities (termed the 'cascade equations' by Julia and Silva [1998]).[71] We state here a simplified version of the Boundary theorem in which the action is left unchanged by an infinitesimal symmetry transformation (i.e. we do not allow for the possibility of a surface term).[72]

The Boundary theorem (restricted form)

If a continuous group of transformations depending smoothly on ρ arbitrary functions of time and space $p_k(x)$ $(k = 1, 2, ..., \rho)$ and their first derivatives is a Noether symmetry group[73] of the Euler-Lagrange equations associated with a Lagrangian $L(\phi_i, \partial_\mu \phi_i, x^\mu)$, then the following three sets of ρ relations are satisfied, one for every parameter on which the symmetry group depends:

$$\sum_i \partial_\mu (b^\mu_{ki} E^L_i) = \partial_\mu j^\mu_k \tag{5}$$

$$\sum_i (b^\mu_{ki} E^L_i) = j^\mu_k - \sum_i \left[\partial_\nu \left(\frac{\partial L}{\partial(\partial_\nu \phi_i)} b^\mu_{ki}\right)\right] \tag{6}$$

$$\left(\frac{\partial L}{\partial(\partial_\mu \phi_i)} b^\nu_{ki}\right) + \left(\frac{\partial L}{\partial(\partial_\nu \phi_i)} b^\mu_{ki}\right) = 0. \tag{7}$$

Once again, the b^ν_{ki} depend on the particular transformations of the fields in question, the details of which need not concern us here. The first identity is

[68] See [Brading and Brown, 2007].

[69] For a discussion of this in the case of general covariance, see [Brading and Brown, 2002].

[70] The Boundary theorem also appears in the work of Hermann Weyl, specialized to the case of his unified field theory (see [Weyl, 1922, 287–289]; the first appearance was in the 1919 third edition), and was published in a non-theory-specific form by Utiyama [1956; 1959].

[71] As with Noether's second theorem, the Boundary theorem is a useful tool in the attempt to demarcate 'true gauge theories' from theories where the local symmetry is a 'mere mathematical artefact', through inspection of the identities that result from the theorem, and through the physical significance — or otherwise — of these identities.

[72] For further details of the Boundary theorem, including the generalization that allows for a surface term, see [Brading and Brown, 2007].

[73] The Boundary theorem is here stated in a restricted form such that the Noether symmetry group must belong to the restricted class of such groups associated with an *invariant* action.

connected to the existence of *superpotentials* associated with local symmetries.[74] The second equation can be used to investigate the relationship between a field and its sources. For example, in the case of classical electromagnetism, we can investigate the relationship between the local gauge symmetry of the theory and the condition that:

$$j^\mu = \partial_\nu F^{\mu\nu}, \tag{8}$$

i.e. that Maxwell's equations with dynamical sources hold. Using the case of classical electromagnetism as our example once again, the third equation becomes the condition that the electromagnetic tensor be antisymmetric (showing the relationship between this condition and the local gauge symmetry of that theory):

$$F^{\mu\nu} + F^{\nu\mu} = 0. \tag{9}$$

These remarks have been necessarily brief, and the reader is referred to [Barbashov and Nesterenko, 1983], along with Brading and Brown [2003; 2007], for detailed derivations and discussion of these results. The identities of the Boundary theorem and of Noether's two theorems are not all independent of one another, and which is most useful depends on the context and the question under consideration. As with Noether's theorems, the Boundary theorem holds independently of the specific details of the dynamical equations, and together they allow us to investigate structural features of our theories that are associated with the symmetry properties of those theories.

8 THE INTERPRETATION OF SYMMETRIES IN CLASSICAL PHYSICS

In what follows, we begin with 'Wigner's hierarchy', which has become the canonical view of the relationship between symmetries, laws and events. We supplement this with a brief discussion of the connection between symmetry and irrelevance, and how this bears on the interpretation of the various symmetries described in Section 4.2, above.

The general interpretation of symmetries in physical theories can adopt a number of complementary approaches. We can ask about the different *roles* that various symmetries play; about the epistemological, ontological or other *status* that various symmetries have; and about the significance of the structures left invariant by symmetry transformations. We end with some remarks on each of these issues.

8.1 Wigner's hierarchy

The starting point for contemporary philosophical discussion of the status and significance of symmetries in physics is Wigner's 1949 paper 'Invariance in Physical Theory', along with his three later papers published in 1964.[75] In these papers,

[74] See, for example, [Trautman, 1962, 179].

[75] Wigner's papers can be found in the collection *Symmetries and Reflections* [Wigner, 1967].

Wigner makes the distinction mentioned above (see Section 4.2) between geometrical and dynamical symmetries, which we will return to below. He also presents his view of the *hierarchy* of physical knowledge, according to which symmetries are viewed as properties of laws:

> There is a strange hierarchy in our knowledge of the world around us. Every moment brings surprises and unforeseeable events — truly the future is uncertain. There is, nevertheless, a structure in the events around us, that is, correlations between the events of which we take cognizance. It is this structure, these correlations, which science wishes to discover, or at least the precise and sharply defined correlations. ... We know many laws of nature and we hope and expect to discover more. Nobody can foresee the next such law that will be discovered. Nevertheless, there is a structure in the laws of nature which we call the laws of invariance. This structure is so far-reaching in some cases that laws of nature were guessed on the basis of the postulate that they fit into the invariance structure. ... This then, the progression from events to laws of nature, and from laws of nature to symmetry or invariance principles, is what I meant by the hierarchy of our knowledge of the world around us. [Wigner, 1967, 28–30].

This view of symmetries, as properties of laws, has become canonical.

8.2 Symmetry and irrelevance

There is a general property of laws, or of the underlying events, to which symmetries are connected: the *irrelevance* of certain quantities that might otherwise be thought to have physical significance.[76] In Section 4.2 we outlined the variety of symmetries found in physics, and in each case the symmetry is associated with a property that is deemed irrelevant for the purposes of describing the law-governed behaviour of a system. For example, left-right symmetry means that whether a system is left-handed or right-handed is irrelevant to its law-governed evolution. Famously, this symmetry is violated in the weak interaction: the law-governed behaviour of systems turns out to be sensitive to handedness for certain processes (see [Pooley, 2003]).

In Section 4.2 we characterized the distinction between global and local symmetries mathematically, in terms of the dependence on constant parameters and arbitrary functions of time (and space) respectively. The physical meaning of this distinction can be understood through the associated properties that are deemed *irrelevant*. A *global* symmetry reflects the irrelevance of absolute values of a certain quantity: only relative values are relevant. So in Newtonian mechanics, for example, spatial translation invariance holds and absolute position is irrelevant to

[76] For an analysis of the connection between symmetry, equivalence and irrelevance, see [Castellani, 2003].

the behaviour of systems.[77] Only relative positions matter, and this is reflected in the structure of the theory through the equations being invariant under global spatial translations — the equations do not depend upon, or invoke, a background structure of absolute positions.

A global symmetry is a special case of a local symmetry. A *local* symmetry reflects the irrelevance not only of absolute values, but furthermore of relative values specified at-a-distance: only *local* relative values (i.e. relative values specified at a point) are relevant. This is reflected in the structure of the theory by the equations of motion not depending upon some background structure that determines relative values at-a-distance (i.e. there is no global background structure associated with the property in question).[78]

8.3 Roles of symmetries

The various different roles in which symmetries are invoked in physics have become much more evident with the advent of quantum theory.[79] Nevertheless, already with the classification of crystals using their remarkable and varied symmetry properties, we see the powerful *classificatory* role at work. Indeed, it was with René-Just Haüy's use of symmetries in this way that crystallography emerged in 1801 as a discipline distinct from mineralogy.[80] Furthermore, the *heuristic* and/or *normative* role is clear for the principle of relativity in the construction of both Special and General Relativity (see above, Section 5). The *unificatory role*, so prominent now in the attempts to unify the fundamental forces, was already present (although differing methodologically somewhat) in Hilbert's attempt to construct a generally covariant theory of gravitation and electromagnetism (see [Sauer, 1999]) and in Weyl's 1918 unified theory of gravitation and electromagnetism, for example. Symmetries may also be invoked in a variety of *explanatory* roles. For example, on the basis of Noether's first theorem (see Section 7) we might say that it is *because* of the translational symmetry of classical mechanics (plus

[77] We are considering here Newton*ian* mechanics, without Newton's absolute space.

[78] Instead, we require the explicit appearance of a *connection* in our theory, which provides the rules by which two distant objects may be brought together so that comparisons between them may be made locally.

[79] The application of the theory of groups and their representations for the exploitation of symmetries in the quantum mechanics of the 1920s represents a dramatic step-change in the significance of symmetries in physics, with respect to both the foundations and the phenomenological interpretation of the theory. As Wigner emphasized on many occasions, one essential reason for the 'increased effectiveness of invariance principles in quantum theory' [Wigner, 1967, 47] is the linear nature of the state space of a quantum physical system, corresponding to the possibility of superposing quantum states. For details on the application of symmetries in quantum physics we refer the reader to [Dickson, this vol., ch. 4, Section 3.3], [Landsman, this vol., ch. 5, Section 4.1], and [Halvorson, this vol., ch. 8, Section 5.2]. For philosophical discussions see [Brading and Castellani, 2003].

[80] The use of discrete symmetries in crystallography continued through the nineteenth century in the work of J. F. Hessel and A. Bravais, leading to the 32 point transformation crystal classes and the 14 Bravais lattices. These were combined into the 230 space groups in the 1890s by E. S. Fedorov, A. Schönflies, and W. Barlow. The theory of discrete groups continues to be important in such fields as solid state physics, chemistry, and materials science.

satisfaction of other conditions) that linear momentum is conserved in that theory. Another example would be an appeal to symmetry principles as an explanation, via Wigner's hierarchy, for (i) aspects of the *form* of the laws, and thereby (ii) why certain events occur and others do not.

8.4 Status of symmetries

Are symmetries ontological, epistemological, or methodological in status? It is clear that symmetries have an important heuristic function, as discussed above (Section 5) in the context of relativity. This indicates a methodological status, something that becomes further developed within the context of quantum theory. We can also ask whether we should attribute an ontological or epistemological status to symmetries.

According to an ontological viewpoint, symmetries are seen as "existing in nature", or characterizing the structure of the physical world. One reason for attributing symmetries to nature is the so-called geometrical interpretation of spatiotemporal symmetries, according to which the spatiotemporal symmetries *of physical laws* are interpreted as symmetries *of spacetime itself*, the "geometrical structure" of the physical world. Moreover, this way of seeing symmetries can be extended to non-external symmetries, by considering them as properties of other kinds of spaces, usually known as "internal spaces".[81] The question of exactly what a realist would be committed to on such a view of internal spaces remains open, and an interesting topic for discussion.

One approach to investigating the limits of an ontological stance with respect to symmetries would be to investigate their empirical or observational status: can the symmetries in question be directly observed? We first have to address what it means for a symmetry to be observable, and indeed whether all symmetries have the same observational status. Kosso [2000] arrives at the conclusion that there are important differences in the empirical status of the different kinds of symmetries. In particular, while global continuous symmetries can be directly observed — via such experiments as the Galilean ship experiment — a local continuous symmetry can have only indirect empirical evidence.[82]

The direct observational status of the familiar global spacetime symmetries leads us to an epistemological aspect of symmetries. According to Wigner, the spatiotemporal invariance principles play the role of a prerequisite for the very possibility of discovering the laws of nature: 'if the correlations between events changed from day to day, and would be different for different points of space, it would be impossible to discover them' [Wigner, 1967]. For Wigner, this conception of symmetry principles is essentially related to our ignorance (if we could directly know all the laws of nature, we would not need to use symmetry principles in our search for them). Such a view might be given a methodological interpretation, ac-

[81] See Section 4.2, above, for the varieties of symmetry.

[82] See Section 6.1, above; and Brading and Brown [2003b], who argue for a different interpretation of Kosso's examples.

cording to which such spatiotemporal regularities are presupposed in order for the enterprize of discovering the laws of physics to get off the ground.[83] Others have arrived at a view according to which symmetry principles function as "transcendental principles" in the Kantian sense (see for instance [Mainzer, 1996]). It should be noted in this regard that Wigner's starting point, as quoted above, does not imply exact symmetries — all that is needed epistemologically (or methodologically) is that the global symmetries hold approximately, for suitable spatiotemporal regions, so that there is sufficient stability and regularity in the events for the laws of nature to be discovered.

As this discussion, and that of the preceding Subsections, indicate, the *differences* between various types of symmetry become important before we have ventured very far into interpretational issues. For this reason, much recent work on the interpretation of symmetry in physical theory has focussed not on general questions, such as those sketched above, but on addressing interpretational questions specific to particular symmetries.[84]

8.5 Symmetries, objectivity, and objects

Turning now to the issue of the structures left invariant by symmetry transformations, the old and natural idea that what is objective should not depend upon the particular perspective under which it is taken into consideration is reformulated in the following group theoretical terms: what is objective is what is invariant with respect to the relevant transformation group. This connection between symmetries and objectivity is something that has a long history going back to the early twentieth century at least. It was highlighted by Weyl [1952], where he writes that 'We found that objectivity means invariance with respect to the group of automorphisms.' This connection between objectivity and invariance was discussed particularly in the context of Relativity Theory, both Special and General. We recall Minkowski's famous phrase ([1908] 1923, 75) that 'Henceforth space by itself, and time by itself, are doomed to fade away into mere shadows, and only a kind of union of the two will preserve an independent reality', following his geometrization of Einstein's Special Theory of Relativity, and the recognition of the spacetime interval (rather than intervals of space and of time) as the geometrically invariant quantity. The connection between objectivity and invariance in General Relativity was discussed by, amongst others, Hilbert and Weyl, and continues to be an issue today.[85]

[83]We are grateful to Brandon Fogel for this point, and for the comparison he suggested between this view of spatiotemporal symmetries and the methodological face of Einstein's notion of separability.

[84]These include the varieties of gauge invariance found in classical electromagnetism and in quantum theories, along with general covariance in GTR (these being continuous symmetries), plus the discrete symmetries of parity (violated in the weak interaction) and permutation invariance, both of which are found in classical theory but require reconsideration in the light of quantum theory. See [Brading and Castellani, 2003].

[85]We saw above (Sections 6.2 and 6.3) some aspects of this debate in the discussion of Einstein's 'hole argument' and of the status of observables in GTR.

Related to this is the use of symmetries to characterize the *objects* of physics as sets of invariants. Originally developed in the context of quantum theory, this approach can also be applied in classical physics.[86] The basic idea is that the invariant quantities — such as mass and charge — are those by which we characterize objects. Thus, through the application of group theory we can use symmetry considerations to determine the invariant quantities and "construct" or "constitute" objects as sets of these invariants.[87]

In conclusion, then, the philosophical questions associated with symmetries in classical physics are wide-ranging. What we have offered here is nothing more than an overview, influenced by our own interests and puzzles, which we hope will be of service in further explorations of this philosophically and physically rich field.

ACKNOWLEDGEMENTS

We are grateful to the editors, Jeremy Butterfield and John Earman, for their encouragement and detailed comments. We would also like to thank Brandon Fogel, Brian Pitts, and Thomas A. Ryckman for their comments and suggestions.

BIBLIOGRAPHY

[Adler *et al.*, 1975] R. Adler, M. Bazin, and M. Schiffer. *Introduction to General Relativity.* 2nd edition, McGraw-Hill Book Company, New York, St Louis, San Francisco, Toronto, London, Sydney, 1975.

[Anderson, 1967] J. L. Anderson. *Principles of Relativity Physics.* New York and London: Academic press, 1967.

[Barbashov and Nesterenko, 1983] B. M. Barbashov and V. V. Nesterenko. Continuous symmetries in field theory, *Fortschr. Phys.*, 31: 535-67, 1983.

[Barbour, 1989] J. B. Barbour. *Absolute or Relative Motion?* Vol. 1 *The Discovery of Dynamics.* Oxford University Press, 1989.

[Barbour, 2005] J. B. Barbour. *Absolute or Relative Motion?* Vol. 2 *The Deep Structure of General Relativity.* Oxford University Press, 2005.

[Belot, 1998] G. Belot. Understanding electromagnetism. *British Journal for the Philosophy of Science*, 49: 531–55, 1998.

[Belot, 2006] G. Belot. The representation of time and change in mechanics. This volume.

[Brading, 2005] K. Brading. A note on general relativity, energy conservation, and Noether's theorems. In J. Eisenstaedt and A. J. Kox, Birkhäuser (eds.) *The Universe of General Relativity (Einstein Studies)*, 2005.

[Brading and Brown, 2002] K. Brading and H. R. Brown. General covariance from the perspective of Noether's theorems. *Diálogos*: 59-86, 2002.

[Brading and Brown, 2003] K. Brading and H. R. Brown. Symmetries and Noether's theorems. In K. Brading and E. Castellani (eds.), pages 89-109, 2003.

[Brading and Brown, 2004] K. Brading and H. R. Brown. Are gauge symmetry transformations observable?. *British Journal for the Philosophy of Science*, 55: 645-65, 2004.

[Brading and Brown, 2007] K. Brading and H. R. Brown. Noether's theorems, gauge symmetries and general relativity. Manuscript, 2007.

[Brading and Castellani, 2003] K. Brading and E. Castellani (eds.). *Symmetry in Physics: Philosophical Reflections.* Cambridge University Press, 2003.

[86] See Max Born, reprinted in [Castellani, 1998].

[87] For further discussion see [Castellani, 1998, part II].

[Brading and Castellani, 2006] K. Brading and E. Castellani. Curie's Principle, Encore. In preparation, 2006.

[Brading and Ryckman, 2007] K. Brading and T. A. Ryckman. Hilbert's axiomatic method and the foundations of physics: an interpretation of generally covariant physics and a revision of Kant's epistemology. Manuscript, 2007.

[Brown, 2006] H. Brown. *Physical Relativity: Spacetime Structure from a Dynamical Perspective*. Oxford University Press, 2006.

[Butterfield, 2004] J. Butterfield. Between laws and models: Some philosophical morals of Lagrangian mechanics; available at Los Alamos arXive: http://arxiv.org/abs/physics/0409030; and at Pittsburgh archive: http://philsci-archive.pitt.edu/archive/00001937/.

[Butterfield, 2006] J. Butterfield. On symmetry and conserved quantities in classical mechanics. Forthcoming in W. Demopoulos and I. Pitowsky (eds.), *Festschrift for Jeffrey Bub*. Kluwer: University of Western Ontario Series in Philosophy of Science, 2006; available at Los Alamos arXive: http://arxiv.org/abs/physics/; and at Pittsburgh archive: http://philsci-archive.pitt.edu/archive/00002362/.

[Butterfield, 2006] J. Butterfield. On symplectic reduction in classical mechanics. This volume.

[Castellani, 1998] E. Castellani (ed.). *Interpreting Bodies. Classical and Quantum Objects in Modern Physics*. Princeton University Press, 1998.

[Castellani, 2003] E. Castellani. Symmetry and equivalence. In K. Brading and E. Castellani (eds.), pages 321-334, 2003.

[Chalmers, 1970] A. F. Chalmers. Curie's principle. *The British Journal for the Philosophy of Science*, 21: 133-148, 1970.

[Corry, 2004] L. Corry. *David Hilbert and the Axiomatization of Physics (1898-1918)*. Dordrecht: Kluwer Academic, 2004.

[Curie, 1894] P. Curie. Sur la symétrie dans les phénomènes physiques. Symétrie d'un champ électrique et d'un champ magnétique. *Journal de Physique*, 3^e série, vol. 3: 393-417, 1894.

[Curie, 1981] P. Curie. On symmetry in physical phenomena. Trans. J. Rosen and P. Copié, *Am. J. Phys.*, 49(4): 17-25, 1981.

[Descartes, 1998] R. Descartes. *The World and other Writings*. S. Gaukroger (ed. and trans.), Cambridge University Press, 1998.

[Dickson, 2006] M. Dickson. Non-relativistic quantum theory. This volume, 2006.

[Doughty, 1990] N. Doughty. *Lagrangian Interaction*. Addison-Wesley, 1990.

[Earman, 2004] J. Earman. Curie's principle and spontaneous symmetry breaking. *International Studies in Philosophy of Science*, 18: 173-199, 2004.

[Earman, 2006a] J. Earman. Two challenges to the substantive requirement of general covariance. *Synthese*, in press, 2006.

[Earman, 2006b] J. Earman. Aspects of determinism in modern physics. This volume.

[Halvorson, 2006] H. Halvorson. Algebraic quantum field theory. This volume.

[Hawkins, 2000] T. Hawkins. *Emergence of the Theory of Lie Groups: an Essay in the History of Mathematics 1869-1926*, New York: Springer, 2000.

[Hilbert, 1915] D. Hilbert. Die Grundlagen der Physik (Erste Mitteilung), Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen. *Mathematisch-physikalische Klasse*: 395-407, 1915.

['t Hooft , 2006] 't Hooft. The conceptual basis of quantum field theory. This volume.

[Howard, 2007] D. Howard. 'And I shall not mingle conjectures and certainties': The roots of Einstein's principle theories/constructive theories distinction, manuscript, 2007.

[Ismael, 1997] J. Ismael. Curie's principle. *Synthese*, 110: 167-190, 1997.

[Janssen, 2005] M. Janssen. Of pots and holes: Einstein's bumpy road to general relativity. *Ann. Phys.*, 14, Supplement: 58-85, 2005.

[Julia and Silva, 1998] B. Julia and S. Silva. Current and superpotentials in classical gauge invariant theories I. Local results with applications to perfect fluids and general relativity, 1998. gr-qc/9804029 v2.

[Klein, 1918] F. Klein. Über die Differentialgesetze für die Erhaltung von Impuls und Energie in der Einsteinschen Gravitationstheorie. Königliche Gesellschaft der Wissenschaften zu Göttingen. Mathematisch-physikalische Klasse. *Nachrichten*: 469-92, 1918.

[Kosso, 2000] P. Kosso. The empirical status of symmetries in physics. *British Journal for the Philosophy of Science*, 51: 81-98, 2000.

[Lanczos, 1949] C. Lanczos. *The Variational Principles of Mechanics*. University of Toronto Press, [1949], 1962.

[Landsman, 2006] N. P. Landsman. Between classical and quantum. This volume.

[Mainzer, 1996] K. Mainzer. *Symmetries of Nature. A Handbook*. New York: de Gruyter, 1996.

[Malament, 2006] D. Malament. Classical relativity theory. This volume.

[Miller, 1981] A. I. Miller. *Albert Einstein's Special Theory of Relativity*. Addison-Wesley Publishing Company, 1981.

[Minkowski, 1908] H. Minkowski. Space and time. In *The Principle of Relativity. A Collection of Original Memoirs on the Special and General Theory of Relativity*. New York: Dover, pages 75-91, [1908], 1923.

[Misner *et al.*, 1970] C. W. Misner, K. S. Thorne, and J. A. Wheeler. *Gravitation*. New York: W. H. Freeman and Company, 1970.

[Noether, 1971] E. Noether. Noether's theorem. *Transport Theory and Statistical Physics*, 1: 183-207, 1971.

[Norton, 1984] J. Norton. How Einstein found his field equations: 1912-1915. *Hist. Stud. Phys. Sci.*, 14: 253-316, 1984.

[Norton, 1993] J. Norton. General covariance and the foundations of general relativity: eight decades of dispute. *Rep. Prog. Phys.*, 56: 791-858, 1993.

[Norton, 2003] J. Norton. General covariance, gauge theories, and the Kretschmann objection. In K. Brading and E. Castellani (eds.), pages 110-123, 2003.

[Ohanian and Ruffini, 1994] H. C. Ohanian and R. Ruffini. *Gravitation and Spacetime*, 2nd edition. London, New York: W. W. Norton and company, 1994.

[Olver, 1993] P. J. Olver. *Applications of Lie Groups to Differential Equations*, 2nd edition, New York: Springer, 1993.

[O'Raifeartaigh, 1997] L. O'Raifeartaigh. *The Dawning of Gauge Theory*. Princeton University Press, 1997.

[Pitts, 2006] J. B. Pitts. Absolute objects and counterexamples: Jones-Geroch Dust, Torretti constant curvature, Tetrad-Spinor, and scalar density. *Studies in History and Philosophy of Modern Physics*, forthcoming, 2006. Manuscript available at Los Alamos arXive: http://arxiv.org/abs/gr-qc/0506102/.

[Pooley, 2003] O. Pooley. Handedness, parity violation, and the reality of space. In K. Brading and E. Castellani (eds.), pages 250-280, 2003.

[Radicati, 1987] L. A. Radicati. Remarks on the early developments of the notion of symmetry breaking. In M. G. Doncel, A. Hermann, L. Michel, and A. Pais (eds.), *Symmetries in physics (1600-1980)*. Barcelona: Servei de Publicaciones, pages 195-206, 1987.

[Rovelli, 2006] C. Rovelli. Quantum gravity. This volume.

[Ryckman, 2005] T. Ryckman. *The Reign of Relativity*. Oxford University Press, 2005.

[Sauer, 1999] T. Sauer. The relativity of discovery: Hilbert's first note on the foundations of physics. *Arch. Hist. Exact Sci.*, 53: 529-575.

[Shubnikow and Koptsik, 1974] A. V. Shubnikow and V. A. Koptsik. *Symmetry in Science and Art*. London: Plenum Press, 1974.

[Stachel, 1993] J. Stachel. The meaning of general covariance: the hole story. In J. Earman *et al.* (eds.), *Philosophical Problems of the Internal and External World: Essays on the Philosophy of Adolf Grünbaum.*. University of Pittsburgh Press, pages 129-160, 1993.

[Torretti, 1983] R. Torretti. *Relativity and Geometry*. Oxford: Pergamon Press, 1983.

[Trautman, 1962] A. J. Trautman. Conservation laws in general relativity. In L. Witten (ed.), *Gravitation: An Introduction to Current Research*. New York: John Wiley and Sons, pages 169-98, 1962.

[Utiyama, 1956] R. Utiyama. Invariant theoretical interpretation of interaction. *Physical Review*, 101: 1597-607, 1956.

[Utiyama, 1959] R. Utiyama. Theory of invariant variation in the generalized canonical dynamics. *Prog. Theor. Phys. Suppl.*, 9: 19-44, 1959.

[Wald, 1984] R. Wald. *General Relativity*. University of Chicago Press, 1984.

[Weyl, 1922] H. Weyl. *Space, Time, Matter*. Reprinted by Dover (1952), 1922.

[Weyl, 1952] H. Weyl. *Symmetry*. Princeton University Press, 1952.

[Whittaker, 1904] E. T. Whittaker. *A Treatise on the Analytical Dynamics of Particles and Rigid Bodies*. Cambridge University Press, ([1904] 1989).

[Wigner, 1967] E. P. Wigner. *Symmetries and Reflections*. Bloomington: Indiana University Press, 1967.

[Yaglom, 1988] I. M. Yaglom. *Felix Klein and Sophus Lie. Evolution of the Idea of Symmetry in the Nineteenth Century*. Boston-Basel: Birkhäuser, 1988.

This page intentionally left blank

ASPECTS OF DETERMINISM IN MODERN PHYSICS

John Earman

1 INTRODUCTION

The aims of this chapter are to review some aspects of determinism that are familiar to physicists but are little discussed in the philosophical literature and to show how these aspects connect determinism to issues about symmetries in physics, the structure and ontological status of spacetime, predictability, and computability.[1] It will emerge that in some respects determinism is a robust doctrine and is quite hard to kill, while in other respects it is fragile and requires various enabling assumptions to give it a fighting chance. It will also be seen that determinism is far from a dead issue. Whether or not ordinary non-relativistic quantum mechanics (QM) admits a viable deterministic underpinning is still a matter of debate. Less well known is the fact that in some cases QM turns out to be more deterministic than its classical counterpart. Quantum field theory (QFT) assumes determinism, at least at the classical level, in order to construct the field algebra of quantum observables. Determinism is at the heart of the cosmic censorship hypothesis, the most important unsolved issue in classical general relativity theory (GTR). And issues about the nature and status of determinism lie at the heart of key foundation issues in the search for a theory of quantum gravity.

2 PRELIMINARIES

2.1 The metaphysics of determinism

The proposal is to begin by getting a grip on the doctrine of determinism as it was understood pre-GTR and pre-QM, and then subsequently to try to understand how the doctrine has to be adjusted to accommodate these theories. In pre-GTR physics, spacetime serves as a fixed background against which the drama of physics is enacted. In pre-QM physics it was also assumed that there is a set $\mathcal{O}$ of genuine physical magnitudes (a.k.a. "observables") each of which takes a determinate

[1] Recent surveys of determinism are found in Butterfield [1998], Earman [2004a], and Hoefer [2003]. A collection of articles on various aspects of determinism is found in Atmanspacher and Bishop [2002].

Handbook of the Philosophy of Science. Philosophy of Physics
Volume editors: Jeremy Butterfield and John Earman
General editors: Dov M. Gabbay, Paul Thagard and John Woods
© 2007 Elsevier B.V. All rights reserved.

value at every moment of time; call these the *occurrent* magnitudes. Other physical magnitudes may be dispositional in character and may take on determinate values only in appropriate contexts; but it was assumed that these dispositional magnitudes supervene on the nondispositional magnitudes.[2] A *history* H is a map from $\mathbb{R}$ to tuples of values of the basic magnitudes, where for any $t \in \mathbb{R}$ the *state* $H(t)$ gives a snapshot of behavior of the basic magnitudes at time t. The world is *Laplacian deterministic* with respect to $\mathcal{O}$ just in case for any pair of histories H_1, H_2 satisfying the laws of physics, if $H_1(t) = H_2(t)$ for some t, then $H_1(t) = H_2(t)$ for all t.

Several remarks are in order. First, the 't' which appears in the above definition is supposed to be a *global time function*. This notion can be defined in a manner that applies to classical, special relativistic, and general relativistic spacetimes: a global time function is a smooth map $t : \mathcal{M} \to \mathbb{R}$, where $\mathcal{M}$ is the spacetime manifold, such that for any $p, q \in \mathcal{M}$, $t(p) < t(q)$ just in case there is a future directed timelike curve from p to q.[3] In classical spacetimes, all of which possess an absolute (or observer independent) notion of simultaneity, a timelike curve is one which is oblique to the planes of absolute simultaneity. And the levels $t = const$ of a global time function must coincide with the planes of simultaneity; thus, in the classical setting t is determined up to a transformation of the form $t \to t' = t'(t)$. In the relativistic setting a timelike curve is one whose tangent at any point lies inside the light cone at that point. In causally pathological general relativistic spacetimes (e.g. Gödel spacetime — see Section 6.1) there can be no global time function, and the global sense of Laplacian determinism as defined above makes no sense.[4] But if one global time function exists for a relativistic spacetime, then many exist. A poor choice of global time function can lead to the failure of Laplacian determinism on the above definition. Thus, in the relativistic setting, the definition of determinism must be construed as applying to a suitable choice of time function, the nature of which will be clarified below.

Second, the above formulation of determinism assumes a distinction between laws of nature on one hand and initial/boundary conditions on the other. Where this distinction becomes mushy, so does the doctrine of determinism. There is a

---

[2]The general idea of supervenience is that X supervenes on Y iff within the range of possible cases, there is no difference in X without a difference in Y. The strength and type of supervenience depends on what are counted as possible cases. Here the concern is mainly with physical supervenience where the possible cases are those compatible with the laws of physics.

[3]This definition presupposes that the spacetime is temporally orientable and that one of the orientations has been singled out as giving the future direction of time. The first presupposition is satisfied for classical and special relativistic spacetimes. A general relativistic spacetime (see [Malament, this vol.]) may not be temporally orientable, but a covering spacetime always is since temporal orientability fails only if the spacetime is not simply connected. The second presupposition implies that some solution to the problem of the direction of time has been found (see [Uffink, this vol.]).

[4]A necessary and sufficient condition for the existence of a global time function for a relativistic spacetime is *stable causality* which (roughly speaking) requires that there exists a widening of the null cones that does not result in closed timelike curves; for a precise definition, see [Wald, 1984, 198-199]. Not only does Gödel spacetime not admit a global time function, it does not admit any global time slices (i.e. spacelike hypersurfaces without edges).

huge philosophical literature on laws of nature.[5] Since most of it is unilluminating when it comes to understanding the nature and function of laws in the practice of physics, it will be largely ignored here. For present purposes I will simply stipulate that an acceptable account of laws must satisfy the empiricist constraint that the laws supervene on the totality of non-modal, particular facts.[6] Philosophers like to speculate about non-empiricist laws; but such entities, should they exist, would seem to be beyond the ken of science, and as such they are irrelevant for present purposes. I prefer David Lewis' [1973, 72-77] way of fulfilling the empiricist constraint since it connects the account of laws to the practice of physics: the laws of physics are the axioms or postulates that appear in the ideal theory of physics, where the ideal theory is the one that, within the class of true theories, achieves the best balance between simplicity and information content. All of the historical examples we have of candidates for deterministic laws involve a relatively small subset $\mathcal{B} \subset \mathcal{O}$ of basic occurrent magnitudes, the assumption being that the remaining ones supervene on those of $\mathcal{B}$.[7] This is hardly surprising if, as has been claimed, simplicity is a crucial feature of physical laws. Hermann Weyl shared the conviction that simplicity must figure into an account of laws, but he noted that "this circumstance is apt to weaken the metaphysical power of determinism, since it makes the meaning of natural law depend on the fluctuating distinction between simple and complicated functions or classes of functions" [1932, 42]. This is, I think, a consequence that has to be swallowed and digested. Philosophers who are answerable only to their armchairs are free to think otherwise.

Third, it is conceptually possible that the world could be partially deterministic, i.e. deterministic with respect to partial histories defined by the values of magnitudes in some proper subset $\mathcal{D} \subset \mathcal{O}$ of the occurrent physical magnitudes but non-deterministic with respect to partial histories defined by the values of magnitudes in some other proper subset $\mathcal{N} \subset \mathcal{O}$. But it is hard to imagine a scenario in which this could happen if both $\mathcal{D}$ and $\mathcal{N}$ are basic magnitudes. For in order that the non-deterministic evolution of the elements $\mathcal{N}$ not upset deterministic evolution for $\mathcal{D}$, the magnitudes in $\mathcal{N}$ must not interact with those in $\mathcal{D}$, or else there would have to be a conspiracy in which the upsetting effects of the $\mathcal{N}$ magnitudes on $\mathcal{D}$ cancel out, which is operationally the same. However, this plausibility consideration fails to operate when the $\mathcal{N}$ are non-basic magnitudes; in particular, as discussed below, stochastic processes on one level can supervene on deterministic processes at a lower level (see [Butterfield, 1998]). This fact makes the inference from observed stochastic behavior to indeterminism fraught with peril.

Fourth, the laws of physics typically take the form of differential equations, in which case the issue of Laplacian determinism translates into the question of whether the equations admit an *initial value formulation*, i.e. whether for ar-

[5]For an overview of different accounts of laws of nature, see [Carroll, 2004].

[6]This is what David Lewis has termed "Humean supervenience" with regards to laws of nature; for a defense, see [Earman and Roberts, 2006].

[7]For example, in classical particle mechanics the elements of $\mathcal{B}$ are the positions and momenta of the particles, and it is assumed that any other genuine mechanical magnitude can be expressed as a functional of these basic magnitudes.

bitrary initial data there exists a unique solution agreeing with the given initial data.[8] What counts as initial data depends on the details of the case, but typically it consists of the instantaneous values of the independent variables in the equations, together with the instantaneous values of a finite number of time derivatives of these variables. "Arbitrary" initial data might be thought to include any kinematically possible values of the relevant variable — as with the initial values of particle positions and velocities in Newtonian mechanics — but "arbitrary" must be taken to mean arbitrary within the bounds of compatibility with the equations of motion, which may impose non-trivial constraints on the initial data. This leads to the next remark.

Fifth, in the relativistic setting, field equations often factor into constraint equations, which place restrictions on the initial data, and the evolution equations, which govern how initial data satisfying the constraint equations evolve over time — Maxwell's equations for electromagnetism and Einstein's gravitational field equations being prime examples. In these instances the evolution equations guarantee that once the constraint equations are satisfied they continue to be satisfied over time. This should be a feature of deterministic equations, for if the data at some future time in the unique solution picked out by the initial data do not satisfy the constraints, then the laws are self-undermining. It could be argued that a basic feature of time in relativistic worlds — perhaps the key feature that separates the time dimension from the space dimensions — lies precisely in this separation of evolution and constraint equations.[9]

Sixth, while there is no *a priori* guarantee that the laws of the ideal theory of physics will be deterministic, the history of physics shows that determinism is taken to be what might be termed a 'defeasible methodological imperative': start by assuming that determinism is true; if the candidate laws discovered so far are not deterministic, then presume that there are other laws to be discovered, or that the ones so far discovered are only approximations to the correct laws; only after long and repeated failure may we entertain the hypothesis that the failure to find deterministic laws does not represent a lack of imagination or diligence on our part but reflects the fact that Nature is non-deterministic. An expression of this sentiment can be found in the work of Max Planck, one of the founders of quantum physics: determinism (a.k.a. the law of causality), he wrote, is a "heuristic principle, a signpost and in my opinion the most valuable signpost we possess, to guide us through the motley disorder of events and to indicate the direction in which scientific inquiry should proceed in order to attain fruitful results" [1932, 26; my translation].[10]

---

[8] And as will be discussed below, there are further issues, such as whether the solution depends continuously on the initial data.

[9] See [Callender, 2005] and [Skow, 2005] for defenses of related views on the difference between space and time.

[10] For a history of the debates about the status of determinism among the founding fathers of QM, see [Cushing, 1994] and [Stöltzner, 2003].

2.2 Varieties of determinism

There is a tendency in the philosophical literature to fixate on the Laplacian variety of determinism. But other kinds of determinism crop up in physics. For example, some processes are described by delay-differential equations for which instantaneous initial data may not suffice to single out a unique solution. A simple example is given by the first order ordinary differential equation (ode) $\dot{x}(t) = x(t - C)$ with a constant delay $C > 0$. Laplacian determinism fails since given initial data $x(0)$ is compatible with multiple solutions. However, a near cousin of Laplacian determinism holds since a specification of $x(t)$ for the interval of time $t \in [-C, 0]$ fixes a unique solution.[11] If the constant delay C replaced by a function $\tau(t)$ of t which is unbounded, or if the delay-differential equation has a more complicated form than in the simple example, then even the weakened forms of Laplacian determinism can fail. An illustration of the latter is given by the equation of motion is $\dot{x}(t) = f(t)x(t - 1)$ where $f(t)$ is a continuous function that vanishes outside of $[0, 1]$ and satisfies $\int f(t)dt = -1$. Raju [1994, 120ff] gives an example of an f such that unless $x(t)$ is identically 0 for all $t \geq 1$, the equation of motion admits no solutions for $t < 0$; whereas if $x(t)$ is identically zero for $t \geq 1$, then the equation of motion admits an infinity of solutions for $t < 0$. Changing the delay term $x(t - 1)$ in this example to an advance term $x(t + 1)$ produces an example where an entire past history fails to fix a unique future solution. Very little is known about the initial value problem for what is probably the most important physical application of delay/advance differential equations; namely, charged particles moving under their mutual retarded/advanced interactions.[12]

For sake of definiteness, fix on the Laplacian variety of determinism. Within this variety there is a distinction between future and past determinism. *Past Laplacian determinism* means that for any pair of histories H_1, H_2 satisfying the laws of physics, if $H_1(t) = H_2(t)$ for some t, then $H_1(t') = H_2(t')$ for all $t' > t$. *Future Laplacian determinism* is defined analogously. In principle, Laplacian determinism can hold in one direction of time but not in the other. However, if the laws of motion are time reversal invariant, then future and past determinism stand or fall together. Time reversal invariance is the property that if H is a history satisfying the laws, then so is the 'time reverse' history H^T, where $H^T(t) := {}^R H(-t)$ and where 'R' is the reversal operation that is defined on a case-by-case basis, usually by analogy with classical particle mechanics where $H(t) = (\mathbf{x}(t), \mathbf{p}(t))$, with $\mathbf{x}(t)$ and $\mathbf{p}(t)$ being specifications respectively of the particle positions and momenta at t, and ${}^R H(t) = (\mathbf{x}(t), -\mathbf{p}(t))$.[13] Since all of the plausible candidates for fundamental laws of physics, save those for the weak interactions

---

[11] See [Driver, 1977] for relevant results concerning delay-differential equations.

[12] Driver [1979] studied the special case of identically charged particles confined to move symmetrically on the x-axis under half-retarded and half-advanced interactions. He showed that, provided the particles are sufficiently far apart when they come to rest, a unique solution is determined by their positions when they come to rest.

[13] A different account of time reversal invariance is given in [Albert, 2000, Ch. 1]; but see [Earman, 2002] and [Malament, 2004].

of elementary particles, are time reversal invariant, the distinction between past and future determinism is often ignored.

This is the first hint that there are interesting connections between determinism and symmetry properties.[14] Many other examples will be encountered below, starting with the following section.

2.3 Determinism and symmetries: Curie's Principle

The statement of what is now called 'Curie's Principle' was announced in 1894 by Pierre Curie:

> (CP) When certain effects show a certain asymmetry, this asymmetry must be found in the causes which gave rise to it. [Curie 1894, 401]

Some commentators see in this Principle profound truth, while others see only falsity, and still others see triviality (compare [Chalmers, 1970]; [Radicati, 1987]; [van Fraassen, 1991, 23–24], and [Ismael, 1997]). My reading of (CP) makes it a necessary truth. It takes (CP) to assert a conditional:

If

> (CP1) the laws of motion governing the system are deterministic; and (CP2) the laws of motion governing the system are invariant under a symmetry transformation; and (CP3) the initial state of the system is invariant under said symmetry

then

> (CP4) the final state of the system is also invariant under said symmetry

When the first clause (CP1) in the antecedent holds, the second clause (CP2) can be understood as follows: if an initial state is evolved for a period Δt and then the said symmetry is applied to the (unique) evolved state, the result is the same as first applying the symmetry to the initial state and evolving the resulting state for a period Δt. With this understanding, the reader can easily derive (CP4) from (CP1)-(CP3). Concrete instantiations of Curie's principle at work in classical and relativistic physics can be found in [Earman, 2004b]. An instantiation for GTR is mentioned in Section 6.3 below.[15]

Although (CP) is a necessary truth, it is far from a triviality since it helps to guide the search for a causal explanation of an asymmetry in what is regarded as the final state of system: either the asymmetry is already present in the initial state; or else the initial state is symmetric and the asymmetry creeps in over time,

[14]See [Brading and Castellani, this vol.] for a discussion of symmetries and invariances in modern physics.

[15]For additional remarks on Curie's principle, see [Brading and Castellani, this vol.].

either because the laws that govern the evolution of the system do not respect the symmetry or because they are non-deterministic. If, as is often the case, the latter two possibilities are ruled out, then the asymmetry in the final state must be traceable to an asymmetry in the initial state. It is also worth noting that the use of (CP) has ramifications for the never ending debate over scientific realism; for the asymmetry in the initial state may be imperceptible not only to the naked eye but to any macroscopic means of detection.[16]

3 DETERMINISM AND INDETERMINISM IN CLASSICAL PHYSICS

3.1 The hard road to determinism in classical physics

Classical physics is widely assumed to provide a friendly environment for determinism. In fact, determinism must overcome a number of obstacles in order to achieve success in this setting. First, classical spacetime structure may not be sufficiently rich to support Laplacian determinism for particle motions. Second, even if the spacetime structure is rich, uniqueness can fail in the initial value problem for Newtonian equations of motion if the force function does not satisfy suitable continuity conditions. Third, the equations of motion that typically arise for classical particles plus classical fields, or for classical fields alone, do not admit an initial value formulation unless supplementary conditions are imposed. Fourth, even in cases where local (in time) uniqueness holds for the initial value problem, solutions can break down after a finite time.

The following subsection takes up the first of these topics — the connection between determinism and the structure and ontology of classical spacetimes. The others are taken up in due course.

3.2 Determinism, spacetime structure, and spacetime ontology

Here is the (naive) reason for thinking that neither Laplacian determinism nor any of its cousins stands a chance unless supported by enough spacetime structure of the right kind. Assume that the (fixed) classical spacetime background is characterized by a differentiable manifold $\mathcal{M}$ and various geometric object fields $O_1, O_2, ..., O_M$ on $\mathcal{M}$. And assume that the laws of physics take the form of equations whose variables are the O_i's and additional object fields $P_1, P_2, ..., P_N$ describing the physical contents of the spacetime. (For the sake of concreteness, the reader might want to think of the case where the P_j's are vector fields whose integral curves are supposed to be the world lines of particles.) A symmetry of the spacetime is a diffeomorphism d of $\mathcal{M}$ onto itself which preserves the background structure given by the O_i's — symbolically, $d^*O_i = O_i$ for all values if i, where

[16]For a more detailed discussion of Curie's Principle and its connection to spontaneous symmetry breaking in quantum field theory see [Earman, 2004b]; for spontaneous symmetry breaking in quantum statistical physics, see [Emch, this vol.].

d^* denotes the drag along by d.[17] By the assumption on the form of the laws, a spacetime symmetry d must also be a symmetry of the laws of motion in the sense that if $\langle \mathcal{M}, O_1, O_2, ..., O_M, P_1, P_2, ..., P_N \rangle$ satisfies the laws of motion, then so does $\langle \mathcal{M}, O_1, O_2, ..., O_M, d^*P_1, d^*P_2, ..., d^*P_N \rangle$.[18]

Now the poorer the structure of the background spacetime, the richer the spacetime symmetries. And if the spacetime symmetry group is sufficiently rich, it will contain elements that are the identity map on the portion of spacetime on or below some time slice $t = const$ but non-identity above. We can call such a map a 'determinism killing symmetry' because when applied to any solution of the equations of motion, it produces another solution that is the same as the first for all past times but is different from the first at future times, which is a violation of even the weakest version of future Laplacian determinism.

As an example, take *Leibnizian spacetime*,[19] whose structure consists of all and only the following: a notion of absolute or observer-independent simultaneity; a temporal metric (giving the lapse of time between non-simultaneous events); and a Euclidean spatial metric (giving the spatial distance between events lying on a given plane of absolute simultaneity). In a coordinate system (x^α, t), $\alpha = 1, 2, 3$ adapted to this structure, the spacetime symmetries are

$$(1) \quad \begin{array}{lcll} x^\alpha & \rightarrow & x'^\alpha = R^\alpha_\beta(t) x^\beta + a^\alpha(t) & \qquad \alpha, \beta = 1, 2, 3 \\ t & \rightarrow & t' = t + const & \end{array}$$

where $R^\alpha_\beta(t)$ is an orthogonal time dependent matrix and the $a^\alpha(t)$ are arbitrary smooth functions of t. Clearly, the symmetries (1) contain determinism killing symmetries.

It is also worth noting that if the structure of spacetime becomes very minimal, no interesting laws of motion, deterministic or not, seem possible. For example, suppose that the time metric and the space metric are stripped from Leibnizian spacetime, leaving only the planes of absolute simultaneity. And suppose that the laws of physics specify that the world is filled with a plenum of constant mass dust particles and that the world lines of these particles are smooth curves that never cross. Then either every smooth, non-crossing motion of the dust is allowed by the laws of motion or none is, for any two such motions are connected by a symmetry of this minimal spacetime.

Two different strategies for saving determinism in the face of the above construction can be tried. They correspond to radically different attitudes towards

---

[17] A diffeomorphism d of the manifold $\mathcal{M}$ is a one-one mapping of $\mathcal{M}$ onto itself that preserves $\mathcal{M}$'s differentiable structure. For the sake of concreteness, assume that d is C^∞.

[18] For on the assumption that the laws are (say) differential equations relating the O_i and P_j, they cannot be sensitive to the "bare identity" of the points of $\mathcal{M}$ at which the O_i and P_j take some given values. This diffeomorphism invariance of the laws is one of the ingredients of what is called substantive general covariance (see section 6.2). One might contemplate breaking diffeomorphism invariance by introducing names for individual spacetime points; but the occurrence of such names would violate the "universal" character that laws are supposed to have.

[19] The details of various classical spacetime structures are to be found in [Earman, 1989].

the ontology of spacetime. The first strategy is to beef up the structure of the background spacetime. Adding a standard of rotation kills the time dependence in $R^\alpha_\beta(t)$, producing what is called *Maxwellian spacetime*. But since the $a^\alpha(t)$ are still arbitrary functions of t there remain determinism killing symmetries. Adding a standard of inertial or straight line motion linearizes the $a^\alpha(t)$ to $v^\alpha t + c^\alpha$, where the v^α and c^α are constants, producing *neo-Newtonian spacetime*[20] whose symmetries are given by the familiar Galilean transformations

$$(2) \quad \begin{array}{lcll} x^\alpha & \to & x'^\alpha = R^\alpha_\beta x^\beta + v^\alpha t + c^\alpha & \qquad \alpha, \beta = 1, 2, 3. \\ t & \to & t' = t + const & \end{array}$$

The mappings indicated by (2) do not contain determinism killing symmetries since if such a map is the identity map for a finite stretch of time, no matter how short, then it is the identity map period. Note that this way of saving determinism carries with it an allegiance to "absolute" quantities of motion: in neo-Newtonian spacetime it makes good sense to ask whether an isolated particle is accelerating or whether an isolated extended body is rotating. To be sure, this absolute acceleration and rotation can be called 'relational' quantities, but the second place in the relation is provided by the structure of the spacetime — in particular, by the inertial structure — and not by other material bodies, as is contemplated by those who champion relational accounts of motion.

The second strategy for saving determinism proceeds not by beefing up the structure of the background spacetime but by attacking a hidden assumption of the above construction — the "container view" of spacetime. Picturesquely, this assumption amounts to thinking of spacetime as a medium in which particles and fields reside. More precisely, in terms of the above apparatus, it amounts to the assumption that $\langle \mathcal{M}, O_1, O_2, ..., O_M, P_1, P_2, ..., P_N \rangle$ and $\langle \mathcal{M}, O_1, O_2, ..., O_M, d^*P_1, d^*P_2, ..., d^*P_N \rangle$, where d is any diffeomorphism of $\mathcal{M}$ such that $d^*P_j \neq P_j$ for some j, describe different physical situations, even when d is a spacetime symmetry, i.e. $d^*O_i = O_i$ for all i. Rejecting the container view leads to (one form of) relationism about spacetime. A spacetime relationist will take the above construction to show that, on pain of abandoning the possibility of determinism, those who are relationists about motion should also be relationists about spacetime. Relationists about motion hold that talk of absolute motion is nonsensical and that all meaningful talk about motion must be construed as talk about the relative motions of material bodies. They are, thus, unable to avail themselves of the beef-up strategy for saving determinism; so, if they want determinism, they must grasp the lifeline of relationism about spacetime.

Relationalism about motion is a venerable position, but historically it has been characterized more by promises than performances. Newton produced a stunningly successful theory of the motions of terrestrial and celestial bodies. Newton's opponents promised that they could produce theories just as empirically adequate

[20]Full Newtonian spacetime adds a distinguished inertial frame — 'absolute space' — thus killing the velocity term in (2).

and as explanatorily powerful as his without resorting to the absolute quantities of motion he postulated. But mainly what they produced was bluster rather than workable theories.[21] Only in the twentieth century were such theories constructed (see [Barbour, 1974] and [Barbour and Bertotti, 1977]; and see [Barbour, 1999] for the historical antecedents of these theories), well after Einstein's GTR swept away the notion of a fixed background spacetime and radically altered the terms of the absolute vs. relational debate.

3.3 Determinism and gauge symmetries

When philosophers hear the word "gauge" they think of elementary particle physics, Yang-Mills theories, etc. This is a myopic view. Examples of non-trivial gauge freedom arise even in classical physics — in fact, we just encountered an example in the preceding subsection. The gauge notion arises for a theory where there is "surplus structure" (to use Michael Redhead's phrase) in the sense that the state descriptions provided by the theory correspond many-one to physical states. For such a theory a gauge transformation is, by definition, a transformation that connects those descriptions that correspond to the same physical state.

The history of physics shows that the primary reason for seeing gauge freedom at work is to maintain determinism. This thesis has solid support for the class of cases of most relevance to modern physics, viz. where the equations of motion/field equations are derivable from an action principle and, thus, the equations of motion are in the form of Euler-Lagrange equations.[22] When the Lagrangian is non-singular, the appropriate initial data picks out a unique solution of the Euler-Lagrange equations and Laplacian determinism holds.[23] If, however, the action admits as variational symmetries a Lie group whose parameters are arbitrary functions of the independent variables, then we have a case of underdetermination because Noether's second theorem tells us that the Euler-Lagrange equations have to satisfy a set of mathematical identities.[24] When these independent variables include time, arbitrary functions of time will show up in solutions to the Euler-Lagrange equations, apparently wrecking determinism.

The point can be illustrated with the help of a humble example of particle mechanics constructed within the Maxwellian spacetime introduced in the preceding subsection. An appropriate Lagrangian invariant under the symmetries of this spacetime is given by

$$\text{(3)} \quad L = \sum \sum\nolimits_{j<k} \frac{m_j m_k}{2M}(\dot{\mathbf{x}}_j - \dot{\mathbf{x}}_k)^2 - V(|\mathbf{x}_j - \mathbf{x}_k|), \qquad M := \sum_i m_i.$$

[21] Newton's opponents were correct in one respect: Newton's postulation of absolute space, in the sense of a distinguished inertial frame was not needed to support his laws of motion.

[22] See [Butterfield, this vol.] and [Belot, this vol.] for accounts of the Lagrangian and Hamiltonian formalisms.

[23] At least if the continuity assumptions discussed in Section 3.5 below are imposed.

[24] For an account of the Noether theorems, see [Brading and Brown, 2003] and [Brading and Castellani, this vol.].

This Lagrangian is singular in the sense that Hessian matrix $\partial^2 L/\partial\dot{\mathbf{x}}_i\partial\dot{\mathbf{x}}_j$ does not have an inverse. The Euler-Lagrange equations are

$$(4) \quad \frac{d}{dt}\left(m_i(\dot{\mathbf{x}}_j - \frac{1}{M}\sum_k m_k\dot{\mathbf{x}}_k)\right) = \frac{\partial V}{\partial \dot{\mathbf{x}}_i}.$$

These equations do not determine the evolution of the particle positions uniquely: if $\mathbf{x}_i(t)$ is a solution, so is $\mathbf{x}'_i(t) = \mathbf{x}_i(t) + \mathbf{f}(t)$, for arbitrary $\mathbf{f}(t)$, confirming the intuitive argument given above for the apparent breakdown of determinism. Determinism can be restored by taking the transformation $\mathbf{x}_i(t) \to \mathbf{x}_i(t) + \mathbf{f}(t)$ as a gauge transformation.

The systematic development of this approach to gauge was carried out by P. A. M. Dirac in the context of the Hamiltonian formalism.[25] A singular Lagrangian system corresponds to a constrained Hamiltonian system. The *primary constraints* appear as a result of the definition of the canonical momenta. (In the simple case of a first-order Lagrangian $L(q, \dot{q}, t)$, where q stands for the configuration variables and $\dot{q} := dq/dt$, the canonical momentum is $p := \partial L/\partial \dot{q}$.) The *secondary constraints* arise as a consequence of the demand that the primary constraints be preserved by the motion. The total set of constraints picks out the *constraint surface* $\mathcal{C}(q,p)$ of the Hamiltonian phase space $\Gamma(q,p)$. The *first class constraints* are those that commute on $\mathcal{C}(q,p)$ with all of the constraints. It is these first class constraints that are taken as the generators of the gauge transformations. The gauge invariant quantities (a.k.a. "observables") are then the phase function $F : \Gamma(q,p) \to \mathbb{R}$ that are constant along the gauge orbits.

Applying the formalism to our toy case of particle mechanics in Maxwellian spacetime, the canonical momenta are:

$$(5) \quad \mathbf{p}_i := \frac{\partial L}{\partial \dot{\mathbf{x}}_i} = \frac{m_i}{M}\sum_k m_k(\dot{\mathbf{x}}_i - \dot{\mathbf{x}}_k) = m_i\dot{\mathbf{x}}_i - \frac{m_i}{M}\sum_k m_k\dot{\mathbf{x}}_k.$$

These momenta are not independent but must satisfy three primary constraints, which require the vanishing of the x, y, and z-components of the total momentum:

$$(6) \quad \phi_\alpha = \sum_i p_i^\alpha = 0, \quad \alpha = 1, 2, 3.$$

These primary constraints are the only constraints — there are no secondary constraints — and they are all first class. These constraints generate in each configuration variable $\mathbf{x}_i$ the same gauge freedom; namely, a Euclidean shift given by the same arbitrary function of time. The gauge invariant variables, such relative particle positions and relative particle momenta, do evolve deterministically.

The technical elaboration of the constraint formalism is complicated, but one should not lose sight of the fact that the desire to save determinism is the motivation driving the enterprise. Here is a relevant passage from [Henneaux and

[25] The standard reference on these matters is [Henneaux and Teitelboim, 1992]. For a user friendly treatment of this formalism, see [Earman, 2003].

Teitelboim, 1992], one of the standard references on constrained Hamiltonian systems:

> The presence of arbitrary functions ... in the total Hamiltonian tells us that not all the q's and p's [the configuration variables and their canonical momenta] are observable [i.e. genuine physical magnitudes]. In other words, although the physical state is uniquely defined once a set of q's and p's is given, the converse is not true — i.e., there is more than one set of values of the canonical variables representing a given physical state. To see how this conclusion comes about, we note that if we are given an initial set of canonical variables at the time t_1 and thereby completely define the physical state at that time, we expect the equations of motion to *fully determine the physical state at other times*. Thus, by definition, any ambiguity in the value of the canonical variables at $t_2 \neq t_1$ should be a physically irrelevant ambiguity. [pp. 16–17]

As suggested by the quotation, the standard reaction to the apparent failure of determinism is to blame the appearance on the redundancy of the descriptive apparatus: the correspondence between the state descriptions in terms of the original variables — the q's and p's — and the physical state is many-to-one; when this descriptive redundancy is removed, the physical state is seen to evolve deterministically. There may be technical difficulties is carrying through this reaction. For example, attempting to produce a reduced phase space — whose state descriptions corresponding one-one to physical states — by quotienting out the gauge orbits can result in singularities. But when such technical obstructions are not met, normal (i.e. unconstrained) Hamiltonian dynamics applies to the reduced phase space, and the reduced phase space variables evolve deterministically.

In addition to this standard reaction to the apparent failure of determinism in the above examples, two others are possible. The first heterodoxy takes the apparent violation of determinism to be genuine. This amounts to (a) treating what the constraint formalism counts as gauge dependent quantities as genuine physical magnitudes, and (b) denying that these magnitudes are governed by laws which, when conjoined with the laws already in play, restore determinism. The second heterodoxy accepts the orthodox conclusion that the apparent failure of determinism is merely apparent; but it departs from orthodoxy by accepting (a), and it departs from the first heterodoxy by denying (b) and, accordingly, postulates the existence of additional laws that restore determinism. Instances that superficially conform to part (a) of the two heterodoxies are easy to construct from examples found in physics texts where the initial value problem is solved by supplementing the equations of motion, stated in terms of gauge-dependent variables, with a gauge condition that fixes a unique solution. For instance, Maxwell's equations written in terms of electromagnetic potentials do not determine a unique solution corresponding to the initial values of the potentials and their time derivatives. Imposing the Lorentz gauge condition converts Maxwell's equations to second order hyperbolic partial differential equations (pdes) that do admit an initial value formulation (see Section 4.2).[26] Similar examples can be concocted in general relativity theory where orthodoxy treats the metric potentials as gauge variables (see Section 6.2). In these examples orthodoxy is aiming to get at the values of the gauge independent variables via a choice of gauge. If this aim is not kept clearly in mind, the procedure creates the illusion that gauge-dependent variables have physical significance. It is exactly this illusion that the two heterodoxies take as real. The second heterodoxy amounts to taking the gauge conditions not as matters of calculational convenience but as additional physical laws. I know of no historical examples where this heterodoxy has led to fruitful developments in physics.

Since there is no *a priori* guarantee that determinism is true, the fact that the orthodox reading of the constraint formalism guarantees that the equations of motion admit an initial value formulation must mean that substantive assumptions that favor determinism are built into the formalism. That is indeed the case, for the Lagrangian/Hamiltonian formalism imposes a structure on the space of solutions: in the geometric language explained in Chapters 1 and 2 of this volume, the space of solutions has a symplectic or pre-symplectic structure. This formalism certainly is not guaranteed to be applicable to all of the equations of motion the Creator might have chosen as laws of motion; indeed, it is not even guaranteed to be applicable to all Newtonian type second order odes. In the 1880s Helmholtz found a set of necessary conditions for equations of this type to be derivable from an action principle; these conditions were later proved to be (locally) sufficient as well as necessary. After more than a century, the problem of finding necessary and sufficient conditions for more general types of equations of motion, whether in the form of odes or pdes, to be derivable from an action principle is still an active research topic.[27]

3.4 Determinism for fields and fluids in Newtonian physics

Newtonian gravitational theory can be construed as a field theory. The gravitational force is given by $\mathbf{F}_{grav} = -\nabla\varphi$, where the gravitational potential φ satisfies the Poisson equation

(7) $$\nabla^2\varphi = \rho$$

with ρ being the mass density. If φ is a solution to Poisson's equation, then so is $\varphi' = \varphi + g(\mathbf{x})f(t)$ where $g(\mathbf{x})$ is a linear function of the spatial variables and $f(t)$

[26]Where $\mathbf{A}$ is the vector potential and Φ is the scalar potential, the Lorentz gauge requires that

$$\nabla \cdot \mathbf{A} + \frac{\partial\Phi}{\partial t} = 0$$

(with the velocity of light set to unity).

[27]Mathematicians discuss this issue under the heading of the "inverse problem." For precise formulations of the problem and surveys of results, see [Anderson and Thompson, 1992] and [Prince, 2000].

is an arbitrary function of t. Choose f so that $f(t) = 0$ for $t \leq 0$ but $f(t) > 0$ for $t > 0$. The extra gravitational force, proportional to $f(t)$, that a test particle experiences in the primed solution after $t = 0$ is undetermined by anything in the past.

The determinism wrecking solutions to (7) can be ruled out by demanding that gravitational forces be tied to sources. But to dismiss homogeneous solutions to the Poisson equation is to move in the direction of treating the Newtonian gravitational field as a mere mathematical device that is useful in describing gravitational interactions which, at base, are really direct particle interactions.[28] In this way determinism helps to settle the ontology of Newtonian physics: the insistence on determinism in Newtonian physics demotes fields to second-class status. In relativistic physics fields come into their own, and one of the reasons is that the relativistic spacetime structure supports field equations that guarantee deterministic evolution of the fields (see Section 4.2).

In the Newtonian setting the field equations that naturally arise are elliptic (e.g. the Poisson equation) or parabolic, and neither type supports determinism-without-crutches. An example of the latter type of equation is the classical heat equation

$$(8) \quad \nabla^2 \Phi = \kappa \frac{\partial \Phi}{\partial t}$$

where Φ is the temperature variable and κ is the coefficient of heat conductivity.[29] Solutions to (8) can cease to exist after a finite time because the temperature "blows up." Uniqueness also fails since, using the fact that the heat equation propagates heat arbitrarily fast, it is possible to construct surprise solutions Φ_s with the properties that (i) Φ_s is infinitely differentiable, and (ii) $\Phi_s(\mathbf{x}, t) = 0$ for all $t \leq 0$ but $\Phi_s(\mathbf{x}, t) \neq 0$ for $t > 0$ (see [John, 1982, Sec. 7.1]). Because (8) is linear, if Φ is a solution then so is $\Phi' = \Phi + \Phi_s$. And since Φ and Φ' agree for all $t \leq 0$ but differ for $t > 0$, the existence of the surprise solutions completely wrecks determinism.

Uniqueness of solution to (8) can be restored by adding the requirement that $\Phi \geq 0$, as befits its intended interpretation of Φ as temperature; for Widder [1975, 157] has shown that if a solution of $\Phi(\mathbf{x}, t)$ of (8) vanishes at $t = 0$ and is non-negative for all $\mathbf{x}$ and all $t \geq 0$, then it must be identically zero. But one could have wished that, rather than having to use a stipulation of non-negativity to shore up determinism, determinism could be established and then used to show that if the temperature distribution at $t = 0$ is non-negative for all $\mathbf{x}$, then the uniquely determined evolution keeps the temperature non-negative. Alternatively, both uniqueness and existence of solutions of (8) can be obtained by limiting the

[28] This demotion of the status of the Newtonian gravitational field can also be supported by the fact that, unlike the fields that will be encountered in relativistic theories, it carries no energy or momentum.

[29] The fact that this equation is not Galilean invariant need cause no concern since Φ implicitly refers to the temperature of a medium whose rest frame is the preferred frame for describing heat diffusion.

growth of $|\Phi(\mathbf{x},t)|$ as $|\mathbf{x}| \to \infty$. But again one could have wished that such limits on growth could be derived as a consequence of the deterministic evolution rather than having to be stipulated as conditions that enable determinism.

Appearances of begging the question in favor of determinism could be avoided by providing at the outset a clear distinction between kinematics and dynamics, the former being a specification of the space of possible states. For example, a limit on the growth of quantum mechanical wave functions does not beg the question of determinism provided by the Schrödinger equation since the limit follows from the condition that the wave function is an element of a Hilbert space, which is part of the kinematical prescription of QM (see Section 5). Since this prescription is concocted to underwrite the probability interpretation of the wave function, we get the ironic result that the introduction of probabilities, which seems to doom determinism, also serves to support it. The example immediately above, as well as the examples of the preceding subsection and the one at the beginning of this subsection, indicate that in classical physics the kinematical/dynamical distinction can sometimes be relatively fluid and that considerations of determinism are used in deciding where to draw the line. The following example will reinforce this moral.[30]

The Navier-Stokes equations for an incompressible fluid moving in $\mathbb{R}^N$, $N = 2, 3$, read

$$\frac{D\mathbf{u}}{dt} = -\nabla p + \upsilon \Delta \mathbf{u} \tag{9a}$$

$$div(\mathbf{u}) = 0 \tag{9b}$$

where $\mathbf{u}(\mathbf{x},t) = (u^1, u^2, ..., u^N)$ is the velocity of the fluid, $p(\mathbf{x},t)$ is the pressure, $\upsilon = const. \geq 0$ is the coefficient of viscosity, and $D/dt := \partial/\partial t + \sum_{j=1}^{N} u^j \partial/\partial x^j$ is the convective derivative (see Foias at al. 2001 for a comprehensive survey). If the fluid is subject to an external force, an extra term has to be added to the right hand side of (9a). The Euler equations are the special case where $\upsilon = 0$. The initial value problem for (9a-b) is posed by giving the initial data

$$(9) \quad \mathbf{u}(\mathbf{x},0) = \mathbf{u}_0(\mathbf{x})$$

[30] Another reaction to the problems of determinism posed by (8) is to postpone them on the grounds that (8) is merely a phenomenological equation; heat is molecular motion and, thus, the fate of determinism ultimately rests with the character of laws of particle motion. It will be seen below, however, that in order to guarantee determinism for particle motion the helping hand of the stipulation of boundary conditions at infinity is sometimes needed. In any case, the postponement strategy taken to its logical conclusion would mean that no judgment about determinism would be forthcoming until we are in possession of the final theory-of-everything. It seems a better strategy to do today the philosophy of today's physics while recognizing, of course, that today's best theory may be superseded by a better future theory that delivers a different message about determinism.

where $\mathbf{u}_0(\mathbf{x})$ is a smooth (C^∞) divergence-free vector field, and is solved by smooth functions $\mathbf{u}, p \in C^\infty(\mathbb{R}^N x[0,\infty))$ satisfying (9)-(10). For physically reasonable solutions it is required both that $\mathbf{u}_0(\mathbf{x})$ should not grow too large as $|\mathbf{x}| \to \infty$ and that the energy of the fluid is bounded for all time:

$$(10) \quad \int_{\mathbb{R}^N} |\mathbf{u}(\mathbf{x},t)|^2 dx < \infty \text{ for all } t > 0.$$

When $\upsilon = 0$ the energy is conserved, whereas for $\upsilon > 0$ it dissipates.

For $N = 2$ it is known that a physically reasonable smooth solution exists for any given $\mathbf{u}_0(\mathbf{x})$. For $N = 3$ the problem is open. However, for this case it is known that the problem has a positive solution if the time interval $[0,\infty)$ for which the solution is required to exist is replaced by $[0,T)$ where T is a possibly finite number that depends on $\mathbf{u}_0(\mathbf{x})$. When T is finite it is known as the "blowup time" since $|\mathbf{u}(\mathbf{x},t)|$ must become unbounded as t approaches T. For the Euler equations a finite blowup time implies that the vorticity (i.e. the *curl* of $\mathbf{u}(\mathbf{x},t)$) becomes unbounded as t approaches T.

Smooth solutions to the Navier-Stokes equations, when they exist, are known to be unique. This claim would seem to be belied by the symmetries of the Navier-Stokes equations since if $\mathbf{u}(\mathbf{x},t) = \mathbf{f}(\mathbf{x},t)$, $p(\mathbf{x},t) = g(\mathbf{x},t)$ is a solution then so is the transformed $\tilde{\mathbf{u}}(\mathbf{x},t) = \mathbf{f}(\mathbf{x} - \varepsilon\alpha(t),t) + \varepsilon\alpha_t$, $\tilde{p}(\mathbf{x},t) = g(\mathbf{x} - \varepsilon\alpha(t),t) - \varepsilon\mathbf{x} \cdot \alpha_t + \frac{1}{2}\varepsilon^2\alpha_{tt}$, where $\alpha(t)$ is an arbitrary smooth function of t alone (see Olver 1993, pp. 130 and 177 (Exer. 2.15)). Choosing $\alpha(t)$ such that $\alpha(0) = \alpha_t(0) = \alpha_{tt}(0) = 0$ but $\alpha(t) \neq 0$ for $t > 0$ results in different solutions for the same initial data unless $\mathbf{f}(\mathbf{x} - \varepsilon\alpha(t),t) + \varepsilon\alpha_t = \mathbf{f}(\mathbf{x},t)$. However, the transformed solution violates the finiteness of energy condition (11).

The situation on the existence of solutions can be improved as follows. Multiplying (9a-b) by a smooth test function and integrating by parts over $\mathbf{x}$ and t produces integral equations that are well-defined for any $\mathbf{u}(\mathbf{x},t)$ and $p(\mathbf{x},t)$ that are respectively L^2 (square integrable) and L^1 (integrable). Such a pair is called a *weak solution* if it satisfies the integral equations for all test functions. Moving from smooth to weak solutions permits the proof of the existence of a solution for all time. But the move reopens the issue of uniqueness, for the uniqueness of weak solutions for the Navier-Stokes equations is not settled. A striking non-uniqueness result for weak solutions of the Euler equations comes from the construction by Scheffer [1994] and Shnirelman [1997] of self-exciting/self-destroying weak solutions: $\mathbf{u}(\mathbf{x},t) \equiv 0$ for $t < -1$ and $t > 1$, but is non-zero between these times in a compact region of $\mathbb{R}^3$.

It is remarkable that basic questions about determinism for classical equations of motion remain unsettled and that these questions turn on issues that mathematicians regard as worthy of attention. Settling the existence question for smooth solutions for the Navier-Stokes equations in the case of $N = 3$ brings a \$1 million award from the Clay Mathematics Institute (see [Fefferman, 2000]).

3.5 Continuity issues

Consider a single particle of mass m moving on the real line $\mathbb{R}$ in a potential $V(x)$, $x \in \mathbb{R}$. The standard existence and uniqueness theorems for the initial value problem of odes can be used to show that the Newtonian equation of motion

$$(11)\quad m\ddot{x} = F(x) := -\frac{dV}{dx}$$

has a locally (in time) unique solution if the force function $F(x)$ satisfies a Lipschitz condition.[31] An example of a potential that violates the Lipschitz condition at the origin is $-\frac{9}{2}|x|^{4/3}$. For the initial data $x(0) = 0 = \dot{x}(0)$ there are multiple solutions of (12): $x(t) \equiv 0$, $x(t) = t^3$, and $x(t) = -t^3$, where m has been set to unity for convenience. In addition, there are also solutions $x(t)$ where $x(t) = 0$ for $t < k$ and $\pm(t-k)^3$ for $t \geq k$, where k is any positive constant. That such force functions do not turn up in realistic physical situations is an indication that Nature has some respect for determinism. In QM it turns out that Nature can respect determinism while accommodating some of the non-Lipschitz potentials that would wreck Newtonian determinism (see Section 5.2).

3.6 The breakdown of classical solutions

Consider again the case of a single particle of mass m moving on the real line $\mathbb{R}$ in a potential $V(x)$, and suppose that $V(x)$ satisfies the Lipschitz condition, guaranteeing a temporally local unique solution for the initial value problem for the Newtonian equations of motion. However, determinism can fail if the potential is such that the particle is accelerated off to $-\infty$ or $+\infty$ in a finite time.[32] Past determinism is violated because two such solutions can agree for all future times $t \geq t^*$ (say) — no particle is present at these times anywhere in space — but disagree at past times $t < t^*$ on the position and/or velocity of the particle when it is present in space. Since the potential is assumed to be time independent, the equations of motion are time reversal invariant, so taking the time reverses of these escape solutions produces solutions in which hitherto empty space is invaded by particles appearing from spatial infinity. These invader solutions provide violations of future determinism. Piecing together escape and invader solutions produces further insults to determinism.

In the 1890's Paul Painlevé conjectured that for $N > 3$ point mass particles moving in $\mathbb{R}^3$ under their mutually attractive Newtonian gravitational forces, there exist solutions to the Newtonian equations of motion exhibiting non-collision singularities, i.e. although the particles do not collide, the solution ceases to exist

[31] $F(x)$ satisfies the Lipschitz condition in an interval $(a, b) \subset \mathbb{R}$ if there is a constant $K > 0$ such that $|F(x_1) - F(x_2)| \leq K|x_1 - x_2|$ for all $x_1, x_2 \in (a, b)$. A sufficient condition for this is that dF/dx exists, is continuous, and $|dF/dx| \leq K$ on (a, b) for some $K > 0$.

[32] See [Reed and Simon, 1975, Theorem X.5] for necessary and sufficient conditions for this to happen.

after a finite time. Hugo von Zeipel [1908] showed that in such a solution the particle positions must become unbounded in a finite time. Finally, near the close of the 20th century Xia [1992] proved Painlevé conjecture by showing that for $N = 5$ point mass particles, the Newtonian equations of motion admit solutions in which the particles do not collide but nevertheless manage to accelerate themselves off to spatial infinity in a finite time (see [Saari and Xia, 1995] for an accessible survey).

Determinism can recoup its fortunes by means of the device, already mentioned above, of supplementing the usual initial conditions with boundary conditions at infinity. Or consolation can be taken from two remarks. The first remark is that in the natural phase space measure, the set of initial conditions that lead to Xia type escape solutions has measure zero. But it is unknown whether the same is true of all non-collision singularities. The second remark is that the non-collision singularities result from the unrealistic idealization of point mass particles that can achieve unbounded velocities in a finite time by drawing on an infinitely deep potential well. This remark does not suffice to save determinism when an infinity of finite sized particles are considered, as we will see in the next subsection.

It is interesting to note that for point particles moving under mutually attractive Newtonian gravitational forces, QM cures both the collision[33] and non-collision singularities that can spell the breakdown of classical solutions (see Section 5.2). This is more than a mere mathematical curiosity since it is an important ingredient in the explanation of the existence and stability of the hydrogen atom.

3.7 Infinite collections

Consider a collection of billiard balls confined to move along a straight line in Euclidean space. Suppose that the balls act only by contact, that only binary collisions occur, and that each such collision obeys the classical laws of elastic impact. Surely, the reader will say, such a system is as deterministic as it gets. This is so, *if* the collection is finite. But if the collection is infinite and unbounded velocities are permitted, then determinism fails because even with all of the announced restrictions in place the system can seemingly self-excite itself (see [Lanford, 1974]). Pérez Laraudogoitia [2001] shows how to use such infinite collections to create an analogue of the escape solution of the preceding subsection where all of the particles disappear in a finite amount of time. The time reverse of this scenario is one in which space is initially empty, and then without warning an infinite stream of billiard balls pour in from spatial infinity.

Legislating against unbounded velocities or imposing boundary conditions at infinity does not suffice to restore determinism if the billiard balls can be made arbitrarily small [Pérez Laraudogoitia, 2001]. For then a countably infinite collection of them can be Zeno packed into a finite spatial interval, say $(0, 1]$, by placing the center of the first ball at 1, the second at $1/2$, the third at $1/4$, etc. Assume for ease of illustration that all the balls have equal mass ($\equiv 1$). A unit mass cue

[33] A collision singularity occurs when two or more of the point particles collide and the solution cannot be continued through the collision time.

ball moving with unit speed from right to left collides with the first ball and sends a ripple through the Zeno string that lasts for unit time, at the end of which all of the balls are at rest. The boring history in which all the balls are at rest for all time is, of course, also a solution of the laws of impact. Comparing this boring history with the previous one shows that past Laplacian determinism is violated.[34]

This failure of determinism carries with it a violation of the conservation and energy momentum, albeit in a weak sense; namely, in the inertial frame in which the object balls are initially at rest, the total energy and the total momentum each have different values before and after the collisions start, but in every other inertial frame there is no violation simply because the values are infinite both before and after the collisions.[35] Pérez Laraudogoitia [2005] has shown how to construct scenarios in which there is a strong violation of conservation of energy and momentum in that the violation occurs in every inertial frame.

3.8 Domains of dependence

With some artificiality one of the threats to classical determinism discussed above can be summarized using a concept that will also prove very helpful in comparing the fortunes of determinism in classical physics and in relativistic physics. By a *causal curve* let us understand a (piecewise) smooth curve in spacetime that represents the spacetime trajectory for a physically possible transfer of energy/momentum. Define the *future domain of dependence*, $D^+(S)$, of a spacetime region S as the set of all spacetime points p such that any past directed causal curve with future endpoint at p and no past endpoint intersects S. The *past domain of dependence* $D^-(S)$ of S is defined analogously. And the *total domain of dependence* $D(S)$ is the union $D^+(S) \cup D^-(S)$. If $p \notin D(S)$ then it would seem that the state in region S does not suffice to determine the state at p since there is a possible causal process that passes through p but never registers on S.

Since neither the kinematics nor the dynamics of classical physics place an upper bound on the velocity at which energy/momentum can be transferred, it would seem that in principle any timelike curve — i.e. any (piecewise) smooth curve oblique to the planes of absolute simultaneity — can count as a causal curve, and as a consequence $D(S) = \varnothing$ even when S is taken to be an entire plane of absolute simultaneity. The examples from Sections 3.4, 3.6, and 3.7 show how the "in principle" can be realized by some systems satisfying Newtonian laws of motion.

We have seen that some threats to classical determinism can be met by beefing up the structure of classical spacetime. And so it is with the threat currently under consideration. *Full Newtonian spacetime* is what results from neo-Newtonian

[34]The time reverse of the interesting history starts with all the balls initially at rest, and then subsequently the collection self-excites, sending a ripple of collisions from left to right and ejecting the cue ball. If this self-exciting history is physically possible, then future laplacian determinism is violated. However, it might rejected on the grounds that it violated Newton's first law of motion.

[35]For a comment on how the availability of an infinite amount of momentum/energy renders the indeterminism unsurprising, see [Norton, 1999, 1268].

spacetime by adding absolute space in the form of a distinguished inertial frame ('absolute space'). In this setting the spacetime symmetries are small enough that there are now finite invariant velocities (intuitively, velocities as measured relative to absolute space), and thus laws can be formulated that set a finite upper bound on the absolute velocity of causal propagation. Nor is this move necessarily *ad hoc* as shown, for example, by the fact that the formulation of Maxwell's laws of electromagnetism in a classical spacetime setting evidently requires the services of a distinguished inertial frame, the velocity of light c being the velocity as measured in this frame.

But, as is well known, such a formulation is embarrassed by the undetectability of motion with respect to absolute space. This embarrassment provides a direct (albeit anachronistic) route from classical to relativistic spacetime. Adopting for classical spacetimes the same geometric language used in the special and general theories of relativity (see [Earman, 1989, Ch. 2]), absolute space is represented by a covariantly constant timelike vector field A^a, the integral curves of which are the world lines of the points of absolute space. The space metric is represented by a degenerate second rank contravariant tensor h^{ab}, which together with A^a defines a tensor that is formally a Minkowski metric: $\eta^{ab} := h^{ab} - A^a A^b$. The unobservability of absolute motion means that there is no preferred way to split η^{ab} into an h^{ab} part and a $A^a A^b$ part, suggesting that η^{ab} is physically as well as formally a Lorentz metric. As we will see in Section 4.1, this puts determinism on much firmer ground in that domains of dependence of local or global time slices are non-empty in the spacetime setting of STR.

3.9 Determinism, predictability, and chaos

Laplace's vision of a deterministic universe makes reference to an "intelligence" (which commentators have dubbed 'Laplace's Demon'):

> We ought to regard the present state of the universe as the effect of its antecedent state and as the cause of the state that is to follow. An intelligence knowing all of the forces acting in nature at a given instant, as well as the momentary positions of all things in the universe, would be able to comprehend in one single formula the motions of the largest bodies as well as the lightest atoms in the world, provided that its intellect were sufficiently powerful to subject all data to analysis; to it nothing would be uncertain, the future as well as the past would be present to its eyes.[36]

[36] [Laplace, 1820]. English translation from [Nagel, 1961, 281-282]. More than a century earlier Leibniz espoused a similar view: "[O]ne sees then that everything proceeds mathematically — that is, infallibly — in the whole wide world, so that if someone could have sufficient insight into the inner parts of things, and in addition has remembrance and intelligence enough to consider all the circumstances and to take them into account, he would be a prophet and would see the future in the present as in a mirror." Quoted from [Cassirer, 1956, 12].

Perhaps by taking Laplace's vision too literally, philosophers and physicists alike conflate determinism and predictability. The conflation leads them to reason as follows: here is a case where predictability fails; thus, here is a case where determinism fails.[37] This is a mistake that derives from a failure to distinguish determinism — an ontological doctrine about how the world evolves — from predictability — an epistemic doctrine about what can inferred, by various restricted means, about the future (or past) state of the world from a knowledge of its present state.

There is, however, an interesting connection between determinism and practical predictability for laws of motion that admit an initial value problem that is *well-posed* in the sense that, in some appropriate topology, the solutions depend continuously on the initial data.[38] The standard existence and uniqueness proofs for the initial value problem for the odes used in particle mechanics also furnish a proof of well-posedness, which can be traced to the fact that the existence proof is constructive in that it gives a procedure for constructing a series of approximations that converge to the solution determined by the initial data.

To illustrate the implications of well-posedness for predictability, consider the toy case of a system consisting of a single massive particle obeying Newtonian equations of motion. If a suitable Lipschitz condition is satisfied, then for any given values of the position $q(0)$ and velocity $\dot{q}(0)$ of the particle at $t = 0$ there exists (for some finite time interval surrounding $t = 0$) a unique solution: symbolically $q(t) = F(q(0), \dot{q}(0), t)$. And further, since this initial value problem is well-posed, for any fixed $t > 0$ (within the interval for which the solution is guaranteed to exist), F is a continuous function of $q(0)$ and $\dot{q}(0)$. Suppose then that the practical prediction task is to forecast the actual position $\bar{q}(t^*)$ of the particle at some given $t^* > 0$ with an accuracy of $\epsilon > 0$, and suppose that although measurements of position or velocity are not error free, the errors can be made arbitrarily small. By the continuity of F, there exist $\delta_1 > 0$ and $\delta_2 > 0$ such that if $|q(0) - \bar{q}(0)| < \delta_1$ and $|\dot{q}(0) - \bar{\dot{q}}(0)| < \delta_2$, then $|q(t^*) - \bar{q}(t^*)| < \epsilon$. Thus, measuring at $t = 0$ the actual particle position and velocity with accuracies $\pm\delta_1/2$ and $\pm\delta_2/2$ respectively ensures that when the measured values are plugged into F, the value of the function for $t = t^*$ answers to the assigned prediction task. (Note, however, that since the actual initial state is unknown, so are the required accuracies $\pm\delta_1/2$ and $\pm\delta_2/2$, which may depend on the unknown state as well as on ϵ and t^*. This hitch could be overcome if there were minimum but non-zero values of δ_1 and δ_2 that

[37]On the philosophical side, Karl Popper is the prime example. Popper [1982] goes so far as to formulate the doctrine of "scientific determinism" in terms of prediction tasks. An example on the physics side is Reichl [1992]: "[W]e now know that the assumption that Newton's equations are deterministic is a fallacy! Newton's equations are, of course, the correct starting point of mechanics, but in general they only allow us to determine [read: predict] the long time behavior of *integrable* mechanical systems, few of which can be found in nature" (pp. 2–3). I am happy to say that in the second edition of Reichl's book this passage is changed to "[W]e now know that the assumption that Newton's equations can *predict* the future is a fallacy!" [Reichl 2004, 3; italics added].

[38]When the topology is that induced by a norm $||\cdot||$ on the instantaneous states represented by a function $s(t)$ of time, well-posedness requires that there is a non-decreasing, nonnegative function $C(t)$ such that $||s(t)|| \leq C(t)||s(0)||$, $t > 0$, for any solution $s(t)$.

answered to the given prediction task whatever the initial state; but there is no *a priori* guarantee that such minimum values exist. A prior measurement with known accuracy of the position and velocity at some $t^{**} < 0$ will put bounds, which can be calculated from F, on the position and velocity at $t = 0$. And then the minimum values can be calculated for accuracies δ_1 and δ_2 of measurements at $t = 0$ that suffice for the required prediction task for any values of the position and velocity within the calculated bounds.)

Jacques Hadamard, who made seminal contributions to the Cauchy or initial value problem for pdes, took the terminology of "well-posed" (a.k.a. "properly posed") quite literally. For he took it as a criterion for the proper mathematical description of a physical system that the equations of motion admit an initial value formulation in which the solution depends continuously on the initial data (see [Hadamard, 1923, 32]). However, the standard Courant-Hilbert reference work, *Methods of Mathematical Physics*, opines that

> "properly posed" problems are by far not the only ones which appropriately reflect real phenomena. So far, unfortunately, little mathematical progress has been made in the important task of solving or even identifying such problems that are not "properly posed" but still are important and motivated by realistic situations. [1962, Vol. 2, 230].

Some progress can be found in [Payne, 1975] and the references cited therein.

Hadamard was of the opinion that if the time development of a system failed to depend continuously on the initial conditions, then "it would appear to us as being governed by pure chance (which, since Poincaré,[39] has been known to consist precisely in such a discontinuity in determinism) and not obeying any law whatever" [1923, 38]. Currently the opinion is that the appearance of chance in classical systems is due not to the failure of well-posedness but to the presence of chaos.

The introduction of *deterministic chaos* does not change any of the above conclusions about determinism and predictability. There is no generally agreed upon definition of chaos, but the target class of cases can be picked out either in terms of cause or effects. The cause is sensitive dependence of solutions on initial conditions, as indicated, for example, by positive Lyapunov exponents. The effects are various higher order ergodic properties, such as being a mixing system, being a K-system, being a Bernoulli system, etc.[40] Generally a sensitive dependence on initial conditions *plus* compactness of the state space is sufficient to secure such properties. The sensitive dependence of initial condition that is the root cause of chaotic behavior does not contradict the continuous dependence of solutions on initial data, and, therefore, does not undermine the task of predicting with any desired finite accuracy the state at a *fixed* future time, assuming that error in measuring the initial conditions can be made arbitrarily small. If, however, there

[39] See Poincaré's essay "Chance" in *Science and Method* [1952].

[40] See Uffink, this volume, section 6.2, or [Lichtenberg and Lieberman, 1991] for definitions of these concepts.

is a fixed lower bound on the accuracy of measurements — say, because the measuring instruments are macroscopic and cannot make discriminations below some natural macroscopic scale — then the presence of deterministic chaos can make some prediction tasks impossible. In addition, the presence of chaos means that no matter how small the error (if non zero) in ascertaining the initial conditions, the accuracy with which the future state can be forecast degrades rapidly with time. To ensure the ability to predict with some given accuracy $\epsilon > 0$ for all $t > 0$ by ascertaining the initial conditions at $t = 0$ with sufficiently small error $\delta > 0$, it would be necessary to require not only well-posedness but *stability*, which is incompatible with chaos.[41]

Cases of classical chaos also show that determinism on the microlevel is not only compatible with stochastic behavior at the macro level but also that the deterministic microdynamics can ground the macro-stochasticity. For instance, the lowest order ergodic property — ergodicity — arguably justifies the use of the microcanonical probability distribution and provides for a relative frequency interpretation; for it implies that the microcanonical distribution is the only stationary distribution absolutely continuous with respect to Lebesque measure and that the measure of a phase volume is equal to the limiting relative frequency of the time the phase point spends in the volume. In these cases there does not seem to be a valid contrast between "objective" and "epistemic" probabilities. The probabilities are epistemic in the sense that conditionalizing on a mathematically precise knowledge of the initial state reduces the outcome probability to 0 or 1. But the probabilities are not merely epistemic in the sense of merely expressing our ignorance, for they are supervenient on the underlying microdynamics.

Patrick Suppes [1991; 1993] has used such cases to argue that, because we are confined to the macrolevel, determinism becomes for us a "transcendental" issue since we cannot tell whether we are dealing with a case of irreducible stochasticity or a case of deterministic chaos. Although I feel some force to the argument, I am not entirely persuaded. There are two competing hypotheses to explain observed macro-stochasticity: it is due to micro-determinism plus sensitive dependence on initial conditions vs. it is due to irreducible micro-stochasticity. The work in recent decades on deterministic chaos supplies the details on how the first hypothesis can be implemented. The details of the second hypothesis need to be filled in; particular, it has to be explained how the observed macro-stochasticity supervenes on the postulated micro-stochasticity.[42] And then it has to be demonstrated that the two hypotheses are underdetermined by all possible observations on the macrolevel. If both of these demands were met, we would be faced with a particular instance of the general challenge to scientific realism posed by underdetermination of theory by observational evidence, and all of the well-rehearsed moves and countermoves in the realism debate would come into play. But it is futile to fight these battles until some concrete version of the second hypothesis is presented.

[41] Stability with respect to a norm on states $s(t)$ requires that there is a constant C such that $||s(t)|| \leq C||s(0)||$, $t > 0$, for any solution $s(t)$. Compare to footnote 38.

[42] It is not obvious that micro-stochasticity always percolates up to the macro-level.

3.10 Laplacian demons, prediction, and computability

Since we are free to imagine demons with whatever powers we like, let us suppose that Laplace's Demon can ascertain the initial conditions of the system of interest with absolute mathematical precision. As for computational ability, let us suppose that the Demon has at its disposal a universal Turing machine. As impressive as these abilities are, they may not enable the Demon to predict the future state of the system even if it is deterministic. Returning to the example of the Newtonian particle from the preceding subsection, if the values of the position and velocity of the particle at time $t = 0$ are plugged into the function $F(q(0), \dot{q}(0), t)$ that specifies the solution $q(t)$, the result is a function $\mathcal{F}(t)$ of t; and plugging different values of the initial conditions results in different $\mathcal{F}(t)$ — indeed, by the assumption of determinism, the $\mathcal{F}(t)$'s corresponding to different initial conditions must differ on any finite interval of time no matter how small. Since there is a continuum of distinct initial conditions, there is thus a continuum of distinct $\mathcal{F}(t)$'s. But only a countable number of these $\mathcal{F}(t)$'s will be Turing computable functions.[43] Thus, for most of the initial conditions the Demon encounters, it is unable to predict the corresponding particle position $q(t)$ at $t > 0$ by using its universal Turing machine to compute the value of $\mathcal{F}(t)$ at the relevant value of t — in Pitowsky's [1996] happy turn of phrase, the Demon must consult an Oracle in order to make a sure fire prediction.

However, if $q(0)$ and $\dot{q}(0)$ are both Turing computable real numbers, then an Oracle need not be consulted since the corresponding $\mathcal{F}(t)$ is a Turing computable function; and if t is a Turing computable real number, then so is $\mathcal{F}(t)$. This follows from the fact that the existence and uniqueness proofs for odes gives an effective procedure for generating a series of approximations that converges effectively to the solution; hence, if computable initial data are fed into the procedure, the result is an effectively computable solution function. Analogous results need not hold when the equations of motion are pdes. Jumping ahead to the relativistic context, the wave equation for a scalar field provides an example where Turing computability of initial conditions is not preserved by deterministic evolution (see Section 4.4).

A more interesting example where our version of Laplace's Demon must consult an Oracle has been discussed by Moore [1990; 1991] and Pitowsky [1996]. Moore constructed an embedding of an abstract universal Turing machine into a concrete classical mechanical system consisting of a particle bouncing between parabolic and flat mirrors arranged so that the motion of the particle is confined to a unit

---

[43]The familiar notion of a Turing computable or recursive function is formulated for functions of the natural numbers, but it can be generalized so as to apply to functions of the real numbers. First, a computable real number x is defined as a limit of a computable sequence $\{r_n\}$ of rationals that converges effectively, i.e. there is a recursive function $f(n)$ such that $k \geq f(n)$ entails $|x - r_k| \leq 10^n$. Next, a sequence $\{x_n\}$ of reals is said to be computable iff there is a double sequence $\{r_{kn}\}$ such that $r_{kn} \to x_n$ as $k \to \infty$ effectively in both k and n. Finally, a function of the reals is said to be computable iff it maps every computable sequence in its domain into a computable sequence and, moreover, it is effectively uniformly continuous. For details, see [Pour-el and Richards, 1989].

square. Using this embedding Moore was able to show how recursively unsolvable problems can be translated into prediction tasks about the future behavior of the particle that the Demon cannot carry out without help from an Oracle, even if it knows the initial state of the particle with absolute precision! For example, Turing's theorem says that there is no recursive algorithm to decide whether a universal Turing machine halts on a given input. Since the halting state of the universal Turing machine that has been embedded in the particle-mirror system corresponds to the particle's entering a certain region of the unit square to which it is thereafter confined, the Demon cannot predict whether the particle will ever enter this region. The generalization of Turing's theorem by Rice [1953] shows that many questions about the behavior of a universal Turing machine in the unbounded future are recursively unsolvable, and these logical questions will translate into physical questions about the behavior of the particle in the unbounded future that the Demon cannot answer without consulting an Oracle.

The reader might ask why we should fixate on the Turing notion of computability. Why not think of a deterministic mechanical system as an analogue computer, regardless of whether an abstract Turing machine can be embedded in the system? For instance, in the above example of the Newtonian particle with deterministic motion, why not say that the particle is an analogue computer whose motion "computes," for any given initial conditions $q(0), \dot{q}(0)$, the possibly non-Turing computable function $q(t) = F(q(0), \dot{q}(0), t)$? I see nothing wrong with removing the scare quotes and developing a notion of analogue computability along these lines. But the practical value of such a notion is dubious. Determining which function of t is being computed and accessing the value computed for various values of t requires ascertaining the particle position with unbounded accuracy.

Connections between non-Turing computability and general relativistic spacetimes that are inhospitable to a global version of Laplacian determinism will be mentioned below in Section 6.6.

4 DETERMINISM IN SPECIAL RELATIVISTIC PHYSICS

4.1 How the relativistic structure of spacetime improves the fortunes of determinism

Special relativistic theories preserve the Newtonian idea of a fixed spacetime background against which the drama of physics plays itself out, but they replace the background classical spacetimes with Minkowski spacetime. This replacement makes for a tremendous boost in the fortunes of determinism. For the symmetries of Minkowski spacetime are given by the Poincaré group, which admits a finite invariant speed c, the speed of light, making it possible to formulate laws of motion/field equations which satisfy the basic requirement that the symmetries of the spacetime are symmetries of the laws and which propagate energy-momentum no faster than c. For such laws all of the threats to classical determinism that derive from unbounded velocities are swept away.

The last point can be expounded in terms of the apparatus introduced in Section 3.8. For the type of law in question, a causal curve is a spacetime worldline whose tangent at any point lies inside or on the null cone at that point, with the upshot that domains of dependence are now non-trivial. Minkowski spacetime admits a plethora of global time functions. But in contrast with classical spacetimes, such a function t can be chosen so that the domains of dependence $D(t = const)$ of the level surfaces of t are non-empty. Indeed, t can be chosen so that for each and every $t = const$ the domain of dependence $D(t = const)$ is a *Cauchy surface*, i.e. $D(t = const)$ is the entire spacetime. In fact, any inertial time coordinate is an example of a global time function, all of whose levels are Cauchy surfaces.[44] In the context of STR, the definition of Laplacian determinism given above in Section 2.1 is to be understood as applying to a t with this Cauchy property.

It is important to realize that these determinism friendly features just discussed are not automatic consequences of STR itself but involve additional substantive assumptions. The *stress-energy tensor* T^{ab} used in both special and general relativistic physics describes how matter-energy is distributed through spacetime. What is sometimes called the local conservation law for T^{ab}, $\nabla_a T^{ab} = 0$, where ∇_a is the covariant derivative determined by the spacetime metric, does *not* guarantee that the local energy-momentum flow as measured by any observer is always non-spacelike. That guarantee requires also that for any future pointing timelike U^a, $-T^{ab}U_a$ is a future pointing, non-spacelike vector.[45] Combining this requirement with the further demand that the local energy density as measured by any observer is non-negative, i.e. $T^{ab}U_aU_b \geq 0$ for any non-spacelike vector field U^a, produces what is called the *dominant energy condition.* Not surprisingly, this condition, together with the local conservation of T^{ab}, does guarantee that the matter fields that give rise to T^{ab} cannot travel faster than light in the sense that if T^{ab} vanishes on some spacelike region S, then it must also vanish on $D(S)$ (see [Hawking and Ellis, 1973, 91-94]). The dominant energy conditions is thought to be satisfied by all the matter-fields encountered in the actual world, but occasionally what are purported to be counterexamples appear in the physics literature.

4.2 Fundamental fields

In Section 3.4 examples were given to illustrate how fields have a hard time living up to the ideals of Laplacian determinism in classical spacetimes. The situation changes dramatically in Minkowski spacetime, which supports field equations in the form of hyperbolic pdes.[46] For example, the Klein-Gordon equation for a scalar field ϕ of mass $m \geq 0$ obeys the equation

(12) $\nabla_a \nabla^a \phi - m^2 \phi = 0$

[44] Exercise for the reader: Construct a global time function t for Minkowski spacetime such that none of the level surfaces of t are Cauchy.

[45] The minus sign comes from the choice of the signature $(+++-)$ for the spacetime metric.

[46] A standard reference on the classification of pdes relevant to physical applications is [Courant and Hilbert, 1962, Vol. 2]. See also [Beig, 2004].

which is a special case of a linear, diagonal, second order hyperbolic pde. For such equations there is a global existence and uniqueness proof for the initial value problem: given a Cauchy surface Σ of Minkowski spacetime and C^∞ initial data, consisting of the value of ϕ on Σ and the normal derivative of ϕ with respect to Σ, there exists a unique C^∞ solution of (13) throughout spacetime. Furthermore, the initial value problem is well-posed in that (in an appropriate topology) the unique solution depends continuously on the initial data. And finally the Klein-Gordon field propagates causally in that if the initial data are varied outside a closed subset $S \subset \Sigma$, the unique solution on $D(S)$ does not vary. Notice that we have a completely clean example of Laplacian determinism at work — no boundary conditions at infinity or any other enabling measures are needed to fill loopholes through which indeterminism can creep in. By contrast, giving initial data on a timelike hypersurface of Minkowski spacetime is known to lead to an improperly posed Cauchy problem; indeed, not only do solutions not depend continuously on the initial data, but there are C^∞ initial data for which there is no corresponding solution. This asymmetry between the fortunes of determinism in timelike vs. spacelike directions, could, as noted above, be turned around and used as a basis for singling out the time dimension.

It should be emphasized that only restricted classes of hyperbolic pdes are known to have well-posed initial value problems. It is a challenge to mathematical physics to show that the field equations encountered in physical theories can be massaged into a form that belongs to one of these classes. It is a comparatively easy exercise to show that, when written in terms of potentials, the source-free Maxwell equations for the electromagnetic field take the form of a system of linear, diagonal, second order hyperbolic pdes if an appropriate gauge condition is applied to the potentials. In other cases the challenge requires real ingenuity.[47]

Physicists are so convinced of determinism in classical (= non-quantum) special relativistic physics that they sort "fundamental" from "non-fundamental" matter fields according as the field does or does not fulfill Laplacian determinism in the form of global existence and uniqueness theorems for the initial value problem on Minkowski spacetime. The Klein-Gordon field and the source-free Maxwell electromagnetic field qualify as fundamental by this criterion. A dust matter field, however, fails to make the cut since singularities can develop from regular initial data since, for example, in a collapsing ball of dust the density of the dust can become infinite if the outer shells fall inward fast enough that they cross the inner shells. Such shell-crossing singularities can develop even for physically reasonable initial data for the Maxwell-Lorentz equations where the source for the electromagnetic field consists of a charged dust obeying the Lorentz force law. But no great faith in determinism is needed to brush aside the failure of determinism in such examples; they can also be dismissed on the grounds that dust matter is an idealization and, like all idealizations, it ceases to work in some circumstances. Faith in determinism, however, is required to deal with what happens when the Klein-Gordon equation is converted into a non-linear equation by adding terms to

[47]See [Beig, 2004] for examples.

the right hand side of (13), e.g.

$$(13)\quad \nabla_a \nabla^a \phi - m^2 \phi = \lambda \phi^2$$

where λ is a constant. It is known that solutions of (14) corresponding to regular initial data can become infinite at a finite value of t and that such data has non-zero measure (see [Keller, 1957]).

A number of attempts have been made to modify the classical Navier-Stokes equations (see Section 3.4) for dissipative fluids in order to make them consistent with STR in the sense that they become a system of hyperbolic pdes with causal propagation. A criterion of success is typically taken to be that the resulting system admits an initial value formulation, confirming once again the faith in determinism in the special relativistic setting. One difficulty in carrying out this program is that it necessitates the introduction of additional dynamical variables and additional dynamical equations, and as a result many different relativistic generalizations of the classical equations have been produced. Geroch [1995] has argued that we need not be troubled by this *embarras des riches* because the differences among the relativistic generalizations wash out at the level of the empirical observations that are captured by the Navier-Stokes theory.

4.3 Predictability in special relativistic physics

The null cone structure of Minkowski spacetime that makes possible clean examples of Laplacian determinism works against predictability for embodied observers who are not simply "given" initial data but must ferret it out for themselves by causal contact with the system whose future they are attempting to predict. Consider, for example, the predicament of an observer O whose world line is labeled γ in Fig. 1. At spacetime location p this observer decides she wants to predict what will happen to her three days hence (as measured in her proper time).

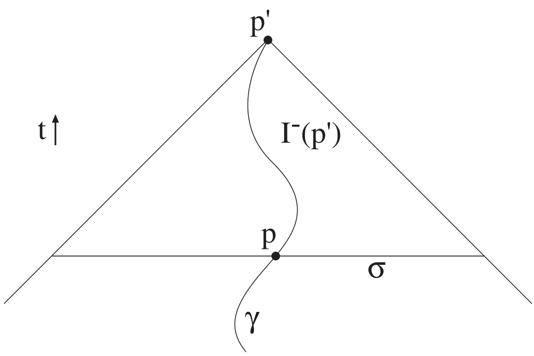


Figure 1. The failure of predictability in Minkowski spacetime

Suppose that, in fact, three-days-hence for O corresponds to the spacetime location p'. And suppose that the equations governing the variables relevant to the prediction are such that initial data on a spacelike hypersurface Σ fixes a unique solution in $D(\Sigma)$. Then to carry out the prediction by way of solving the relevant equations, O must ascertain the state on a local spacelike hypersurface that slices through the past null cone of p', such as σ in Fig. 1. As O's "now" creeps up her world line, the past light of the "now" sweeps out increasingly large portions of σ, but until her "now" reaches p' she does not have causal access to all of σ. And the same goes for any other local slice through the past cone of p'. Thus, the very spacetime structure that provides a secure basis for Laplacian determinism prevents O from acquiring the information she needs before the occurrence of the event that was to be predicted.

This predictability predicament can be formalized in a way that will be useful when it comes to investigating predictability in a general relativistic spacetime $\mathcal{M}, g_{ab}$, where $\mathcal{M}$ is a differentiable manifold and g_{ab} is a Lorentz signature metric defined on all of $\mathcal{M}$, Minkowski spacetime being the special case where $\mathcal{M} = \mathbb{R}^n$ and g_{ab} is the Minkowski metric. Geroch (1977) defines the *domain of prediction* $P(q)$ of a point $q \in \mathcal{M}$ to consist of all points $p \in \mathcal{M}$ such that (i) every past directed timelike curve with future endpoint at p and no past endpoint enters the chronological past $I^-(q)$ of q,[48] and (ii) $I^-(p) \nsubseteq I^-(q)$. Condition (i) is needed to ensure that causal processes that can influence events at p register on the region $I^-(q)$ that is causally accessible to an observer whose "now" is q, and condition (ii) is needed to ensure that from the perspective of q, the events to be predicted at p have not already occurred. The predictability predicament for Minkowski spacetime can now be stated as the theorem that for every point q of Minkowski spacetime, $P(q) = \varnothing$.

Note that the predictability predicament arises not just because of the local null cone structure of Minkowski spacetime but also because of its global topological structure. To drive home this point, suppose that space in (1 + 1)-Minkowski spacetime is compactified to produce the cylindrical spacetime $\mathcal{C}$ pictured in Fig. 2. Now predictability is possible since $I^-(q)$ for any q contains a Cauchy surface, e.g. Σ in Fig. 2. As a result $P(q) = \mathcal{C} - I^-(q)$.

For standard Minkowski spacetime and other spacetimes for which $P(q) = \varnothing$ for every spacetime point q, one can wonder how secure predictions are possible. The answer is that if complete security is required, the only secure predictions have a conditional form, where the antecedent refers to events that are not causally accessible from q. But there will be many such conditionals, with different antecedents and different consequents, and since one will not be in a position to know which of the antecedents is actualized, the best one can do is a "prediction" (all too familiar from economic forecasts) consisting of a big set of conditionals. On the other hand, if complete security is not demanded, then unconditional predictions

[48]For a point q in a relativistic spacetime $\mathcal{M}, g_{ab}$, the chronological past $I^-(q)$ consists of all $p \in \mathcal{M}$ such that there is a future directed timelike curve from p to q. The chronological future $I^+(q)$ of a point q is defined analogously.

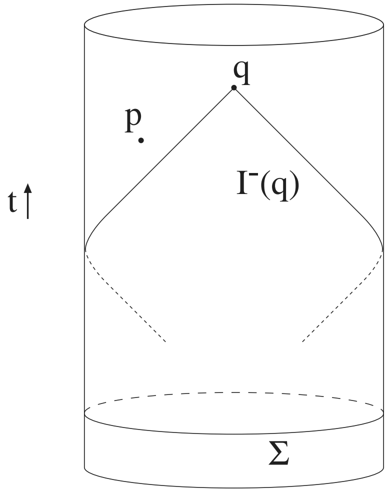


Figure 2. The improved fortunes of predictability when the spatial aspect of Minkowski spacetime is compactified

carrying probability instead of certainty are obtainable if inductive inference from past observations points to one of the antecedents of the set of conditionals as being highly likely.

If one wants predictions that are in principle verifiable, then a third condition needs to be added to the definition of the domain of prediction; namely, (iii) $p \in I^+(q)$. The point p in Fig. 2 satisfies clauses (i) and (ii) but not (iii).

4.4 Special relativity and computability

Pour-el and Richards [1981] constructed an example in which deterministic evolution does not preserve Turing computability. The equation of motion at issue is the relativistic wave equation, which in inertial coordinates is written

$$(14)\quad \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} - \frac{\partial^2 u}{\partial t^2} = 0, \qquad c \equiv 1$$

Pour-el and Richards studied solutions corresponding to initial data at $t = 0$ of the form $u(x, y, z, 0) = f(x, y, z)$, $\partial u(x, y, z, 0)/\partial t = 0$. They showed that there is a Turing computable $f(x, y, z)$ such that the corresponding solution $u(x, y, z, t)$ is not Turing computable at $t = 1$. However, such a solution is necessarily a weak solution (in the sense of Section 3.4) to the wave equation since it must be non-differentiable. And the non-preservation result is sensitive to the norm used to define convergence. Indeed, if Turing computability is defined using the energy norm,[49] then for any Turing computable functions f and g, the solution $u(x, y, z, t)$ corresponding to $u(x, y, z, 0) = f(x, y, z)$, $\partial u(x, y, z, 0)/\partial t = g(x, y, z)$ is Turing computable (see [Pour-el and Richards, 1989, 116-118]).

5 DETERMINISM AND INDETERMINISM IN ORDINARY QM

The folklore on determinism has it that QM is the paradigm example of an indeterministic theory. Like most folklore, this bit contains elements of truth. But like most folklore it ignores important subtleties — in this instance, the fact that in some respects QM is more deterministic and more predictable than classical physics. And to distill the elements of truth from the folklore takes considerable effort — in particular, the folkloric notion that quantum indeterminism arises because the "reduction of the wave packet" is based on a controversial interpretation of the quantum measurement process. Before turning to these matters, I will discuss in Section 5.1 an issue that links to the some of the themes developed above and in Secs. 5.2-5.4 some issues unjustly neglected in the philosophical literature.

5.1 Determinism and Galilean invariance in QM

Here is another example of how linking determinism and symmetry considerations is fruitful in producing physical insights. Consider the motion of a single spinless particle on the real line $\mathbb{R}$, and work in the familiar Hilbert space $\mathcal{H}$ of wave functions, i.e. $\mathcal{H} = L^2_{\mathbb{C}}(\mathbb{R}, dx)$. The evolution of the state $\psi(x) \in \mathcal{H}$ of the quantum particle is governed by the Schrödinger equation

$$(15) \quad i\hbar \frac{\partial \psi}{\partial t} = \hat{H}\psi$$

where $\hat{H}$ is the Hamiltonian operator. This evolution is deterministic, or so it is said. But a puzzle is quickly generated by conjoining the presumed determinism with the presumed Galilean invariance of (16).[50] Since (16) is first order in time, giving the value of the wave function $\psi(x, 0)$ for all $x \in \mathbb{R}$ at $t = 0$ should fix the value of $\psi(x, t)$ for all $t > 0$. But how can this be if the Schrödinger equation is Galilean invariant? A proper Galilean transformation $x \to x' = x - vt$, $v \neq 0$, is the identity map for $t = 0$ but non-identity for $t > 0$. Assuming Galilean invariance of (16), this map must carry a solution to a solution. Since the map in question is the identity for $t = 0$ the two solutions should have the same initial data $\psi(x, 0)$; but since the map is not the identity for $t > 0$ the original solution and

[49] For initial conditions f, g the energy norm is given by
$||f, g||^2 := \int\int\int[(\nabla f)^2 + g^2]dxdydz$.
And for functions u on $\mathbb{R}^4$ the norm is $||u(x, y, z, t)|| = \sup_t E(u, t)$, where
$E(u, t)^2 := \int\int\int[\nabla u + \left(\frac{\partial u}{\partial t}\right)^2]dxdydz$.
If u is a solution of the wave equation, then $E(u, t)$ is independent of t.

[50] For a treatment of the Galilean invariance of the Schrödinger equation, see [Brown, 1999].

its image under a Galilean boost should diverge in the future, violating Laplacian determinism. The resolution of this little puzzle is to reject the implicit assumption that ψ behaves as a scalar under a Galilean transformation. In fact, Galilean invariance of the Schrödinger equation can be shown to imply that the Galilean transformation of ψ depends on the mass of the particle. And this in turn entails a "superselection rule" for mass (discovered by Bargmann [1954]) which means that a superposition of states corresponding to different masses is not physically meaningful in non-relativistic QM.

5.2 How QM can be more deterministic than classical mechanics

Physics textbooks on QM offer a procedure for quantization that starts with a Hamiltonian formulation of the classical dynamics for the system of interest and produces, modulo operator ordering ambiguities, a formal expression for the quantum Hamiltonian operator $\hat{H}$ that is inserted into equation (16).[51] But to make the formal expression into a genuine operator a domain of definition must be specified since, typically, $\hat{H}$ is an unbounded operator and, therefore, is defined at best for a dense domain of the Hilbert space. Usually it is not too difficult to find a dense domain on which $\hat{H}$ acts as a symmetric operator. The question then becomes whether or not this operator is essentially self-adjoint, i.e. has a unique self-adjoint (SA) extension — which will also be denoted by $\hat{H}$.[52] If so, $\hat{U}(t) := \exp(-i\hat{H}t)$ is unitary for all $t \in \mathbb{R}$, and since $\hat{U}(t)$ is defined for the entire Hilbert space, the time evolute $\psi(t) := \hat{U}(t)\psi$ for every vector in the Hilbert space is defined for all times. (The Schrödinger equation (16) is just the "infinitesimal" version of this evolution equation.) Thus, if $\hat{H}$ is essentially SA, none of the problems which beset the deterministic evolution of the classical state can trouble the deterministic evolution of the quantum state.

What is, perhaps, surprising is that the quantum Hamiltonian operator can be essentially SA in some cases where the counterpart classical system does not display deterministic evolution. Recall from Section 3.5 the example of a particle moving on the real line $\mathbb{R}$ in a (real-valued) potential $V(x)$, $x \in \mathbb{R}$. As we saw, when the potential is proportional to $-|x|^{4/3}$ near the origin, the initial value problem for the Newtonian equation of motion does not have a unique solution. But the quantum Hamiltonian operator $\hat{H} = -\dfrac{\hbar^2}{2m}\dfrac{d^2}{dx^2} - V(x)$ is essentially SA provided that $V(x)$ is locally integrable and bounded below. And this can be satisfied by the classically non-deterministic potential by suitably modifying it away from the origin.[53]

---

[51] See [Landsman, this vol.] for details on various approaches to quantization.

[52] A linear operator $\hat{O}$ defined on the dense domain $D(\hat{O})$ of the Hilbert space $\mathcal{H}$ is *symmetric* just in case for all $\psi, \varphi \in D(\hat{O})$, $(\hat{O}\varphi, \psi) = (\varphi, \hat{O}\psi)$, where $(\cdot,\cdot)$ is the inner product on $\mathcal{H}$. That $\hat{O}$ is *self-adjoint* means that $\hat{O} = \hat{O}^*$, i.e. $\hat{O}$ is symmetric and $D(\hat{O}) = D(\hat{O}^*)$, where $\hat{O}^*$ is adjoint of $\hat{O}^*$. Here $D(\hat{O}^*)$ is defined to be the set of $\varphi \in \mathcal{H}$ such that there is a $\chi \in \mathcal{H}$ such that $(\hat{O}\psi, \varphi) = (\psi, \chi)$ for all $\psi \in D(\hat{O})$; then $\hat{O}^*\varphi := \chi$.

[53] An appropriate dense domain is $\{\psi \in L^2_{\mathbb{C}}(\mathbb{R}, dx) : \psi, \psi' \in AC(\mathbb{R})\ \&\ \hat{H}\psi \in L^2_{\mathbb{C}}(\mathbb{R}, dx)\}$ where

Another form of classical indeterminism occurs when the initial value problem has locally in time a unique solution but the solution breaks down after a finite time. The example given in Section 3.6 was that of a system of classical point mass particles moving under the influence of their mutual pairwise attractive $1/r^2$ force, and it was noted that the solution can break down either because of collision or non-collision singularities. Neither type of singularity occurs in the quantum analogue since again the quantum Hamiltonian operator for this case is essentially SA.[54]

QM also cures the indeterminism of the Zeno version of classical billiards discussed in Section 3.7, at least in a backhanded sense. A putative quantum analogue would mimic the Zeno construction of an infinite number of distinct particles in the unit interval $(0, 1]$ by squeezing into that interval an infinite number of wave packets with substantial non-overlap. The latter would require that the uncertainty in position Δx associated with a wave packet becomes arbitrarily small as the origin is approached. By the uncertainty principle, the uncertainty in momentum Δp would have to become unboundedly large as the origin is approached. This represents a breakdown in the putative quantum analogue if Δx and Δp both small in comparison with some specified macroscopic standard is required for mimicing classical behavior.[55]

5.3 How QM (even without state vector reduction) can be a lot less deterministic than classical mechanics

Determinism for the evolution of the quantum state is an all-or-nothing affair. If $\hat{H}$ is essentially SA then the "all" alternative applies since, as already noted, the exponentiation of the unique SA extension gives a unitary evolution operator which is defined for all times and all vectors in the Hilbert space. If $\hat{H}$ is not essentially SA there are two possibilities to consider. The first is that $\hat{H}$ has no SA extensions. This can be dismissed on the grounds that $\hat{H}$ should be a real operator, and every real symmetric operator has SA extensions. The second possibility is that $\hat{H}$ has many SA extensions. Then the "nothing" alternative applies; for the exponentiations of the different SA extensions give physically distinct time evolutions. Roughly speaking, the different self-adjoint extensions correspond to different boundary conditions at the boundary points of the configuration space. Perhaps some boundary condition can be singled out and promoted to lawlike

$AC(\mathbb{R})$ stands for the absolutely continuous functions.

[54]This result is known as Kato's theorem; see [Kato, 1995, Remark 5.6]. For a more detailed discussion of the issue of essential self-adjointness and its implications for quantum determinism, see [Earman, 2005].

[55]Mimicking a classical state in which a particle has given values of position and momentum requires a quantum state ψ that not only returns the given values as expectation values but also gives $(\Delta x)_\psi$ and $(\Delta p)_\psi$ small in comparison with the relevant macroscopic standard; for if $(\Delta x)_\psi$ (respectively, $(\Delta p)_\psi$) is large in comparison with the standard, there is an appreciable probability that the particle will be found with a position (respectively, momentum) different from the given value.

status, thus providing for a unique quantum dynamics. But restoring determinism by this route would require a justification for the hypothesized singling out and promotion. Alternatively, the effects of the non-essential self-adjointness of the Hamiltonian can be down played if it can be shown that the quantum dynamics associated with different self-adjoint extensions all have the same classical limit (see [Earman, 2005b]).

A toy example of the second possibility is given by a particle moving on the positive real line $\mathbb{R}_+$ in a potential $V(x)$, $x \in \mathbb{R}$. If the potential has the form C/x^2, with $C > 0$, then the Newtonian initial value problem has a unique solution, and the solution is defined for all times. The intuitive explanation is that no matter how much energy it has, the classical particle cannot climb over the indefinitely high potential to reach the singularity at the origin, and it cannot reach $x = +\infty$ in finite time. However, the quantum Hamiltonian operator for this case is not essentially SA on $L^2_{\mathbb{C}}(\mathbb{R}_+, dx)$ if $C < \dfrac{3}{4}$ (see [Reed and Simon, 1975, Thm X.10]). The intuitive explanation is that the quantum particle can tunnel through the barrier to reach the singularity, allowing probability to leak away. This leakage is incompatible with unitary evolution, which would obtain as the result of exponentiating the unique SA extension of an essentially SA $\hat{H}$.

The singularity in the configuration space of the toy example is wholly artificial, being created by deleting half of the real line. But analogues in the form of naked or timelike singularities occur in general relativistic spacetimes (see Section 6.4). One can ask whether a relativistic quantum particle propagating on the background of negative mass Schwarzschild spacetime can tunnel through the effective repulsive barrier that surrounds the origin $r = 0$. Horowitz and Marolf [1995] show that the answer is positive.

Essential self-adjointness of the Hamiltonian might be promoted as a selection principle to help decide what systems are "quantum possible," guaranteeing that (barring state vector collapse) the evolution of the quantum is deterministic. Those who think that determinism is an *a priori* truth may look favorably on this promotion, but otherwise its merits are hard to discern.

5.4 Chaos and predictability in QM

QM can not only be more deterministic than classical mechanics, but it can also be more predictable as well. Classical predictability is compromised or completely wrecked for observers who cannot ascertain initial conditions with complete precision if the systems whose behavior they are attempting to predict display sensitive dependence on initial conditions. But if the quantum Hamiltonian operator is essentially SA, then not only is the evolution of the quantum state completely deterministic, its predictability is not compromised by sensitive dependence on initial conditions. The point is simply that the linear, unitary nature of the evolution preserves the Hilbert space norm: $||U(t)\psi_2 - U(t)\psi_1|| = ||U(t)(\psi_2 - \psi_1)|| = ||\psi_2 - \psi_1||$. In words, if two states are close (as measured in the Hilbert space norm) to begin with, they remain close for all times, i.e. the evolution is stable.

This stability causes trouble for anyone seeking chaos in QM itself — they are driven to such extremes as resorting to open systems (whose evolution is not unitary) or to hidden variables whose evolution is not stable.[56] But in itself the stability of quantum evolution poses no *a priori* barrier to explaining how chaos can emerge from quantum dynamics in some appropriate classical limit. For that project only requires showing that the expectation values of relevant classical quantities can diverge rapidly enough (in some appropriate metric) to underwrite higher order ergodic properties that constitute the chaotic behavior observed on the macrolevel (see [Belot and Earman, 1997]). One obvious way to carry out this project is to use Ehrenfest's theorem to show that in the position representation the centroid of a quantum wave packet will follow a classical trajectory, as long as the mean square width of the wave packet remains sufficiently small. However, for classically chaotic trajectories the time interval in which the latter condition holds is uncomfortably short — for example, [Zurek, 1998] estimates that for the chaotic tumbling of Hyperion (a moon of Jupiter) it is of the order of 20 years. Several authors have argued that quantum decoherence comes to the rescue (see, for example, [Zurek, 1998; 2003]), but that is a topic that is beyond the scope of this chapter. Clearly, classical chaos poses a challenge to our understanding of how the classical world emerges from quantum physics.[57] Another aspect of the classical-quantum correspondence is treated in the next section.

5.5 State vector reduction, hidden variables, and all that

Showing that the Hamiltonian operator $\hat{H}$ for a quantum system of interest is essentially SA is not enough to secure the fortunes of determinism for this system, and this for two reasons. The first is that the deterministic evolution of the quantum state might be interrupted by "state vector reduction," as is postulated in some treatments of the measurement problem in QM, by which the unitary evolution $\psi(0) \mapsto \psi(t) = \exp(-i\hat{H}t)\psi(0)$ is suspended and the quantum state jumps into an eigenstate of the observable being measured. In its crudest form state vector reduction is a literal miracle — a violation of the laws of nature — making it an inappropriate topic for the present forum. But there are more sophisticated forms of state vector reduction that model the reduction as a dynamical process. Stochastic models in which the reduction occurs discontinuously and continuously have been studied respectively by Ghirardi *et al.* [1986] and Pearle [1989]. Reduction by means of a non-linear term added to the Schrödinger equation was studied by Pearle [1976]. If the stochastic models of reduction are on the right track and if the stochastic mechanisms they postulate represent irreducible randomness, then obviously determinism is breached. By contrast, the scheme of Pearle [1976]

[56]See the discussions of Kronz [1998] and Cushing and Bowman [1999]. By contrast, physicists who study "quantum chaos" do not try to find chaos in QM itself but rather study the distinguishing properties of quantum systems whose classical counterparts display chaos. For this reason Michael Berry suggested replacing "quantum chaos" with "quantum chaology." Unfortunately the suggestion did not stick.

[57]For a comprehensive survey of this problem, see [Landsman, this vol.].

achieves a deterministic state vector reduction with the help of hidden variables.[58] None of these alternatives to standard non-relativistic QM has been generalized to a viable relativistic quantum field theory, and as far as I am aware none of them play any role in the main lines of research on quantum gravity that come from string theory or loop quantum gravity (see Section 8). Thus, at present it does not seem productive to speculate about the implications for determinism of possible modifications to QM that may or may not become part of some future physics. However, the motivation for introducing state vector reduction is relevant here, for it leads to the second set of reasons why the conventional quantum state dynamics may not be sufficient to secure determinism for the quantum domain.

Classical (= non-quantum) theories wear their interpretations on their sleeves.[59] For example, for a classical theory that admits an (unconstrained) Hamiltonian formulation, observables are in one-one correspondence with functions from the phase space $\Gamma(q,p)$ to (say) the reals $\mathbb{R}$. The intended interpretation is that if f_O is the function corresponding to the observable O, then the value of O at t is o iff the state $(q(t),p(t))$ at t is such that $f_O(q(t),p(t)) = o$. This scheme can be liberalized to allow for dispositional observables which have definite values in only some states; for such an O the representing function f_O is only a partial function. Another liberalization is to allow that the range of f_O includes "fuzzy" (e.g. interval) values. To get an interpretation of QM along similar lines requires adding to the formalism at least two things: (i) an account of which SA operators correspond to quantum observables, and (ii) a semantics for quantum observables in the form of a value assignment rule that specifies what values the observables take under what conditions. I will simply assume that part (i) of the interpretation problem has been solved.

The most obvious way to supply part (ii) would be to ape the classical value assignment rule, replacing the classical state space $\Gamma(q,p)$ by the quantum state space to get a value assignment rule of the form: the value of quantum observable O at t is o iff the state vector $\psi(t)$ is such that $f_O(\psi(t)) = o$ where f_O is the representing function for the quantum observable O. If, as implicitly assumed in this formulation, the quantum state space is taken to be the the unit sphere $\mathbb{S}\mathcal{H}$ of the Hilbert space $\mathcal{H}$ of the system (i.e. $\{\psi \in \mathcal{H} : (\psi,\psi) = 1\}$), then as far as standard QM is concerned, gauge freedom is present since any two elements of $\mathbb{S}\mathcal{H}$ that differ by a phase factor correspond to the same physical state in that all expectation values of observables are the same for the two quantum states. This gauge redundancy can be removed by taking the state space to be the projective Hilbert space $\mathbb{P}\mathcal{H}$, defined as the quotient of $\mathbb{S}\mathcal{H}$ by the action of $\psi \mapsto \varsigma\psi$ where $\varsigma \in \mathbb{C}$ with $|\varsigma| = 1$; equivalently, $\mathbb{P}\mathcal{H}$ is the space of rays or one-dimensional subspaces of $\mathcal{H}$. Thus, from the point of view of conventional QM, the value assignment rule should obey the restriction that $f_O(\psi) = f_O(\psi')$ whenever the

---

[58]The hidden variables are the phase angles, an idea revived by Ax and Kochen [1999]; see below.

[59]But recall from Section 3 that if determinism is demanded, then the initial on-the-sleeve interpretation may have to be modified by seeing gauge freedom at work.

unit vectors ψ' and ψ belong to the same ray. Allowing the value assignment to depend on the phase would amount to introducing "hidden variables," in the terminology used below.

In any case, if the quantum value assignment rule takes the form under discussion and if the problems discussed in Section 5.3 are waived, then arguably QM is a deterministic theory, and this is so even if f_O is a partial function (i.e. is undefined for some quantum states) or even if f_O can take fuzzy values. For assuming no state vector reduction, the state $\psi(0)$ at $t = 0$ uniquely determines the state $\psi(t)$ at any $t > 0$; and assuming the implementation of (ii) under discussion, the state $\psi(0)$ at $t = 0$ uniquely determines the value assignments at any later time $t > 0$. That an observable is assigned no value or a fuzzy value at t does not represent a failure of determinism, which requires only that the laws plus the initial state determine the present and future values of all observables *to the extent that these values are determinate at all.* Thus, on the present option it is a mistake to view the Kochen-Specker theorem, and other subsequent no-go theorems, as showing that QM does not admit a deterministic interpretation. Rather, what these no-go results show is that, subject to certain natural constraints,[60] some subset of quantum observables cannot all be assigned simultaneously sharp values.[61] The same goes for the Bell-type theorems, which are best interpreted as extensions of the Kochen-Specker no-go result to an even smaller set of observables (see [Fine, 1982a; 1982b]).

One example of the type of value assignment rule at issue is the eigenvalue-eigenvector rule which says that, for an observable O whose quantum operator $\hat{O}$ has a discrete spectrum, O has a sharp value at t iff $\hat{O}\psi(t) = o\psi(t)$, in which case $O(t) = o$. But it is just this eigenvalue-eigenvector link that leads to the notorious measurement problem in QM in the form of the inability of the theory to explain why measurements have definite outcomes, and it is this problem that motivated the idea of state vector reduction. In essence the problem arises because of the insistence that "measurement" should not be taken as a primitive term but should be analyzed within QM itself as a physical interaction between the object system and a measuring instrument. But while the application of the standard linear, unitary dynamics to the composite object-system + measurement-apparatus-system can establish a one-one correlation between the eigenstates of the object observable of interest and the eigenstates of the "pointer observable" of the measuring instrument, the application of the eigenvector-eigenvalue rule to the post measurement composite state function yields the unacceptable result that the "pointer" on the measuring instrument is not pointing to any definite value

[60] For example, it is natural to require that if the quantum value assignment rule for O assigns O a definite value, that value lies in the spectrum of the operator $\hat{O}$ corresponding to O. And it is natural to require that for suitable functions g, $F_{g(O)} = g \circ F_O$.

[61] It follows from Gleason's theorem that, subject to very plausible constraints on value assignments, not all of the (uncountably infinite number of) self-adjoint operators in a Hilbert space of dimension 3 or greater can be assigned simultaneously definite values belonging to the spectra of these operators. The Kochen-Specker theorem shows that the same conclusion can be drawn for a finite set of quantum observables. See [Redhead, 1987] for an account of these no-go results.

(see [Albert, 1992] for a detailed exposition). The Schrödinger cat paradox is a cruel illustration of this conundrum in which "live cat" and "dead cat" serve as the "pointer positions."

Thus, if the eigenvalue-eigenvector link is maintained, there are compelling reasons to contemplate a modification of the standard quantum dynamics in order to ensure that in measurement interactions the quantum state evolves into an eigenstate of the relevant observables. But since the decision was made above not to treat such modifications, the discussion that follows will be confined to the other option, namely, the use of a value assignment rule that breaks the eigenvalue-eigenvector link, possibly with the help of "hidden variables" that supplement the quantum state. If hidden variables X are used, the value assignment rule takes the form: the value of quantum observable O at t is o iff the total state $(\psi(t), X(t))$ is such that $f_O(\psi(t), X(t)) = o$, where again f_O stands for the representing function for the observable O but is now a function defined on the augmented state space. If the evolution of the total state is deterministic, then by the same argument as before, the quantum domain is fully deterministic if QM is true. An example is supplied by the Bohm interpretation where $X(t)$ specifies the positions of the particles at t. The quantum component of the total state evolves according to Schrödinger dynamics, and the postulated equation of motion for particle positions guarantees that $(\psi(0), X(0))$ uniquely determines $(\psi(t), X(t))$ for $t > 0$ (see [Holland, 1993] for a survey). On the Bohm interpretation many quantum observables have a dispositional character, taking on determinate values only in adequately specified contexts (typically including a measurement apparatus together with its hidden variables). For example, in the context of a Stern-Gerlach experiment a spin $\frac{1}{2}$ particle will have spin-up (or spin down) just in case the position of the particle lies the appropriate region of the apparatus. The validity of the claim that the Bohm interpretation resolves the measurement problem thus turns on whether all measurements can be reduced to position measurements.

The family of modal interpretations of QM also attempt to resolve the measurement problem by breaking open the eigenvalue-eigenvector link wide enough to allow measurements to have definite outcomes but not so wide as to run afoul of the Kochen-Specker type impossibility results (see [Bub, 1997] and [Dickson, this vol.] for overviews), but in contrast to the Bohm interpretation the modal interpretations have no commitment to maintaining determinism. Very roughly the idea is that an observable associated with a subsystem of a composite system in state $\psi(t)$ has a definite value just in case the reduced density matrix of the subsystem is a weighted sum of projectors associated with an eigenbasis of the observable. This guarantees that in an idealized non-disturbing measurement interaction in which the pointer positions of the measuring instrument are perfectly correlated with the possible values of the object system observable being measured, both the pointer observable and the object system observable have definite values.[62]

[62] More generally, the interaction of a system with its environment will mean that "measurement" of the system is going on all the time. Thus, decoherence aids the modal interpretation by providing the conditions of applicability of the interpretation. In the other direction, deco-

Most forms of the modal interpretation supply the probabilities for an observable to have particular values, assuming that the conditions are appropriate for the observable to have a determinate value; but they are silent as to what the actual value is. Nevertheless, the actual possessed values of quantum observables can be taken to play the role of the hidden variables X, and one can ask whether the total state (ψ, X) can be given a deterministic dynamics. The answer is negative for versions of the modal interpretation discussed in the philosophy literature since these versions do not supply enough hidden variables to allow for determinism. For example, at the time $t > 0$ when an ideal measurement interaction is completed and the eigenstates of pointer position are perfectly correlated with eigenstates of the object observable, the standard modal interpretations say that both the object observable and the pointer observable have definite values. In different runs of the experiment these correlated observables have different values. But in all the runs the initial quantum state $\psi(0)$ is the same, and the experimental situation can be arranged so that modal interpretations say that the initial possessed values $X(0)$ are the same. This failure of determinism is of no concern to the modal theorist whose goal is to solve the measurement problem. To this end it is enough to show that there is a stochastic dynamics for possessed values that is compatible with the statistical predictions of QM. In fact, there is a vast array of such dynamics (see [Dickson, 1997] and [Bacciagaluppi, and Dickson, 1998]).

A different version of the modal interpretation, proposed by Ax and Kochen [1999], takes the option mentioned above of extending the standard quantum state space of rays $\mathbb{P}\mathcal{H}$ to unit vectors $\mathbb{S}\mathcal{H}$. Elements of the former are supposed to characterize statistical ensembles of systems while elements of the latter characterize individual systems. This extension allows the modal interpretation to specify what value an observable has, in circumstances when it has a definite value, and also to provide for a deterministic evolution of the augmented quantum state. It is postulated that the ensemble corresponding to a ray $\varsigma\psi$, $|\varsigma| = 1$, is composed of individual systems with phase factors ς having an initial uniformly random distribution, which accounts for the apparent failure of determinism.

Both the Bohm interpretation and the family of modal interpretations have difficulties coping with relativistic considerations. The former does not have any natural generalization to QFT, at least not one which takes seriously the lesson that in QFT fields are the fundamental entities and particles are epiphenomena of field behavior. The latter does possess a natural generalization to QFT, but it yields the unfortunate consequence that in situations that are standardly discussed, no subsystem observable has a definite value (see [Earman and Ruetsche, 2005] and the references therein).

Many worlds interpretations of QM can be given a literal or a figurative reading (see [Barrett, 1999] for an overview). On the literal reading there are literally many worlds in that spacetime splits into many branches which, from the branch

herence requires something akin to the modal interpretation, for otherwise it does not, contrary to the claims of its promoters, resolve the measurement problem. For more on decoherence, see [Landsman, this vol.].

time onwards, are topologically disconnected from one another (see, for example, [McCall, 1995]).[63] This form of many worlds can be described as a hidden variable interpretation by taking the hidden variables X to describe the spacetime branching and by taking the representing function f_O to be a mapping from the total state $(\psi(t), X(t))$ to a vector, possibly with infinitely many components labeled α, where the component α supplies the value at t of O in branch α. The fate of determinism then depends on whether or not the story of when and how branching takes place makes the evolution of the total state (ψ, X) deterministic. On the figurative reading of "many worlds" there is literally only one world, but there are many minds, contexts, perspectives, or whatever. Also there is no such thing as an observable O simpliciter but rather an observable O-in-context-α, denoted by O_α. If the representing function f_{O_α} is a function of the quantum state only, then determinism seems to be secured. However, our notation is defective in disguising the need for a specification of the contexts that are available at any given time. That specification is determined by the quantum state $\psi(t)$ alone if there is a "democracy of bases," i.e. any "branch" of $\psi(t)$ expressed as a linear combination of the vectors of any orthonormal basis of the Hilbert space of the system defines a context. Such a radical democracy seems incompatible with experience, e.g. in the Schrödinger cat experiment we either see a live cat or we see a dead cat, and we never experience a superposition of seeing a live and seeing a dead cat.[64] To overcome this difficulty some many world theorists propose to work with a preferred set of bases. The issue of determinism then devolves on the question of whether the specification of the set of preferred bases is deterministic. Even if the many worlds interpretation — on either the literal or figurative version — secures ontological determinism, the price seems to be a radical epistemic indeterminism: How do I know which branch of a splitting world or which context of a non-splitting world I am in? Being told that there is no "I" only an "I-in-branch-α" or an "I-in-context-α "is of no help when I — whichever I that is — have to make a prediction about the outcome of a measurement. Here all I can do is fall back on the statistical algorithm of QM. The many worlds interpretation seems to guarantee that even if the world is ontologically deterministic, it behaves, as far as anyone can judge, as if there is an irreducible stochasticity.

Although the discussion of the quantum measurement problem and its ramifications has been very sketchy, I trust it is sufficient to indicate why it is vain to hope for a simple and clean answer to the question of whether the truth of QM entails the falsity of determinism. To arrive at an answer to that question calls for winnowing the various competing interpretations of QM, a task that is far from straightforward, especially since judgments about how to perform the winnowing

[63]How to describe branching spacetimes within the context of conventional spacetime theories is a ticklish matter. Perhaps the most promising move is to hold on to the assumption that spacetime is a differentiable manifold but abandon the assumption that it is a Hausdorff manifold. However, non-Hausdorff manifolds can display various pathological properties that threaten determinism, e.g. geodesics can bifurcate. See Section 6.1.

[64]But how can we be sure? Perhaps momentary mental confusion is a superposition phenomenon.

are inevitably colored by attitudes towards determinism.

6 DETERMINISM IN CLASSICAL GTR

6.1 Einstein's revolution

Einstein's GTR was revolutionary in many respects, but for present purposes the initially most important innovation is that GTR sweeps away the notion — shared by all pre-GTR theories — of a fixed spacetime background against which the evolution of particles and fields takes place. In GTR the spacetime metric is a dynamical field whose evolution is governed by Einstein's gravitational field equations (EFE). Before discussing the issue of whether this evolution is deterministic, two preliminary matters need attention.

First, general relativists typically assume that the manifold $\mathcal{M}$ of a relativistic spacetime $\mathcal{M}, g_{ab}$ is Hausdorff.[65] Without this stipulation determinism would be in deep trouble. For example, non-Hausdorff spacetimes can admit a bifurcating geodesics; that is, there can be smooth mappings γ_1 and γ_2 from, say, $[0, 1]$ into $\mathcal{M}$ such that the image curves $\gamma_1[0, 1]$ and $\gamma_1[0, 1]$ are geodesics that agree for $[0, b)$, $0 < b < 1$, but have different endpoints $\gamma_1(1)$ and $\gamma_2(1)$. According to GTR, the worldline of a massive test particle not acted upon by non-gravitational forces is a timelike geodesic. But how would such a particle know which branch of bifurcating geodesic to follow? Additionally, the (local) uniqueness of solutions to the initial value problem for EFE discussed below in Section 6.3 would fail if non-Hausdorff attachments were allowed.

Second, the reader is reminded that attention is being restricted to relativistic spacetimes $\mathcal{M}, g_{ab}$ that are temporally orientable, and it is assumed that one of the orientations has been singled out as giving a directionality to time. But even with this restriction in place, some of the spacetimes embodied in solutions to EFE are inimical to the formulation of global Laplacian determinism given in Section 2.1. For example, such spacetimes may not admit a global time function. Indeed, the spacetime of the Gödel cosmological model not only does not admit a global time function, but it does not even admit a single global time slice (spacelike hypersurface without edges) so that one cannot meaningfully speak of the universe-at-a-given-moment.[66]

One response would be to narrow down the class of physically acceptable models of GTR by requiring that, in addition to satisfying EFE, such models must also fulfill restrictions on the global causal structure of spacetime that rule out such monstrosities as Gödel's model and other models which contains closed timelike

[65] $\mathcal{M}$ is *Hausdorff* iff for any $p, q \in \mathcal{M}$ with $p \neq q$, there are neighborhoods $N(p)$ and $N(q)$ such that $N(p) \cap N(q) = \emptyset$. Of course, a manifold is (by definition) locally Euclidean and, therefore, locally Hausdorff.

[66] This is a consequence of three features of Gödel spacetime: it is temporally orientable (i.e. it admits a continuous non-vanishing timelike vector field), it is simply connected, and through every spacetime point there passes a closed future directed timelike curve. For a description of the Gödel solution, see [Hawking and Ellis, 1973, 168–170] and [Malament, 1984].

curves. This move has the independent motivation of avoiding the "paradoxes of time travel."[67] But much stronger causality conditions are needed to underwrite the global version of Laplacian determinism in the general relativistic setting.

In the first place, to carry over the classical conception of Laplacian determinism to the context of a general relativistic spacetime requires that the spacetime admit a global time function, which is not guaranteed by the absence of closed timelike curves. But even the requirement of a global time function is not strong enough because it provides no guarantee that the level surfaces of *any* such function will have the Cauchy property. To be at home, Laplacian determinism requires a spacetime $\mathcal{M}, g_{ab}$ that is *globally hyperbolic*, which is the conjunction of two conditions: first, $\mathcal{M}, g_{ab}$ must be *strongly causal* in that for any $p \in \mathcal{M}$ and any neighborhood p there is a subneighborhood such that once a future directed causal curve leaves, it never reenters (intuitively, there are no almost closed causal curves); and second, for every $p, q \in \mathcal{M}$, the causal diamond $J^+(p) \cap J^-(q)$ is compact.[68] Global hyperbolicity guarantees that $\mathcal{M}, g_{ab}$ can be foliated by Cauchy surfaces and that $\mathcal{M}$ is diffeomorphically Σ x $\mathbb{R}$, where Σ is an $n-1$ dimensional manifold if $\dim(\mathcal{M}) = n$. But simply stipulating global hyperbolicity has all the virtues of theft over honest toil. So let us see what can be achieved by honest toil.

6.2 Determinism and gauge freedom in GTR

For pre-relativistic theories a constant theme was that creating an environment friendly to determinism requires willingness to either beef up the structure of the background spacetime or else to see gauge freedom at work in sopping up the apparent indeterminism (recall Section 3.3). But in GTR there is no fixed background structure. Thus, one would expect that GTR either produces indeterminism or else that there is a non-trivial gauge symmetry at work. This expectation is not disappointed.

To see why it is necessary to be more detailed about the EFE:

(16) $$R_{ab} - \frac{1}{2}Rg_{ab} + \Lambda g_{ab} = \kappa T_{ab}$$

where R_{ab} and $R := R^c_c$ are respectively the Ricci tensor (which is defined in terms of g_{ab} and its derivatives) and the Ricci scalar, Λ is the cosmological constant, and T_{ab} is the stress-energy tensor. The cosmological constant can be ignored for present purposes, but it is currently the object of intense interest in cosmology since a positive Λ is one of the candidates for the "dark energy" which is driving the accelerating expansion of the universe (see [Ellis, this vol.]).

A potential model of the theory is then a triple $\langle \mathcal{M}, g_{ab}, T_{ab} \rangle$ where g_{ab}, T_{ab} satisfy (17) at all points of $\mathcal{M}$. Building such a model seems all too easy: start

[67]But see [Earman, Smeenk, and Wüthrich, 2005] which argues that the so-called paradoxes of time travel do not show that time travel is conceptually or physically impossible.

[68]$J^+(p)$ (respectively, $J^-(p)$) denotes the *causal future* (respectively, *causal past*) of p, i.e., the set of all points q such that there is a future directed causal curve from p to q (respectively, from q to p).

with any any general relativistic spacetime $\mathcal{M}, g_{ab}$, compute the Einstein tensor $G_{ab} := R_{ab} - \frac{1}{2}Rg_{ab}$, and define the stress-energy tensor by $T_{ab} := \kappa G_{ab}$. Thus, the understanding must be that T_{ab} arises from a known matter field. And in order to make the environment as friendly as possible for determinism, it will be assumed that the T_{ab}'s that are plugged into the right hand side of (17) fulfill the dominant energy condition (see Section 4.1) which, together with the local conservation law $\nabla^a T_{ab} = 0$ (which itself is a logical consequence of (17)), guarantees that matter-energy does not propagate faster than light.

Even with these enabling stipulations in place, it seems at first glance that determinism gets no traction, at least not if a naively realistic interpretation is given to the models of the theory. The difficulty can be made apparent by repeating a variant of the construction given in Section 3.2. Let $\langle \mathcal{M}, g_{ab}, T_{ab} \rangle$ be a model satisfying all of the above stipulations, and suppose that the spacetime $\mathcal{M}, g_{ab}$ satisfies all of the causality conditions you like, e.g. that it is globally hyperbolic. Since there is no fixed background structure to respect, save for the topological and differentiable structure of $\mathcal{M}$, one is free to choose a diffeomorphism $d : \mathcal{M} \to \mathcal{M}$ such that d is the identity map on and to the past of some Cauchy surface Σ of $\mathcal{M}, g_{ab}$ but non-identity to the future of Σ. Then $\langle \mathcal{M}, d^* g_{ab}, d^* T_{ab} \rangle$, where d^* indicates the drag along by d, will also be a model satisfying all of the stipulations imposed on $\langle \mathcal{M}, g_{ab}, T_{ab} \rangle$. By construction, $d^* g_{ab}(p) = g_{ab}(p)$ and $d^* T_{ab}(p) = T_{ab}(p)$ for all p on or to the past of Σ, but $d^* g_{ab}(p) \neq g_{ab}(p)$ and $d^* T_{ab}(p) \neq T_{ab}(p)$ for some points p to the future of Σ (unless we have inadvertently chosen a d that is a symmetry of g_{ab} and T_{ab}, which can always be avoided). The existence of this pair of models that agree for all past times but disagree in the future is a violation of even the weakest cousin of Laplacian determinism, at least if the surface structure of the theory — tensor fields on a manifold — is taken at face value.

When this sort of construction threatened to undermine determinism in a pre-GTR setting, two options were available for shoring up determinism: add more structure to the background spacetime or reject the container view of spacetime. The first option is ineffective unless the additional elements of spacetime structure are non-dynamical objects, but this represents a retreat from one of the key features Einstein's revolution. If there is to be no retreat, then the second option must be exercised. In the present context the option of rejecting the container view of spacetime takes the form of rejecting the naive realism that reads the theory as describing tensor fields living on a manifold.

Choosing the second option has a principled motivation which is not invented to save determinism in GTR but which follows in line with the treatment of gauge symmetries in pre-general relativistic theories. The field equations (17) of GTR are the Euler-Lagrange equations derived from an action principle that admits the diffeomorphism group as a variational symmetry group. Thus, Noether's second theorem applies, indicating that we have a case of underdetermination — more "unknowns" than there are independent field equations — and that arbitrary functions of the spacetime variables will show up in solutions to the field equations.

Switching from the Lagrangian to the Hamiltonian formulation, it is found, as expected, that GTR is a constrained Hamiltonian theory. There are two families of first class constraints, the momentum constraints and the Hamiltonian constraints.[69] Unfortunately the Poisson bracket algebra of these constraints is not a Lie algebra,[70] and consequently one cannot directly associate the diffeomorphism group, which acts on the spacetime, with a group which acts on the Hamiltonian phase space by finding a natural homomorphism of the Lie algebra of the diffeomorphism group into the constraint algebra. This glitch is overcome by Isham and Kuchař [1986a; 1986b] who show that if appropriate embedding variables and their conjugate momenta are used to enlarge the phase space, then the enlarged constraint algebra is a Lie algebra, and that there exists a homomorphism of the Lie algebra of the spacetime diffeomorphism group into the new constraint algebra. Thus, the standard apparatus for treating gauge symmetries can be applied, yielding the verdict that the diffeomorphism invariance of GTR is to be interpreted as a gauge symmetry. On this interpretation, the above construction does not demonstrate that GTR is indeterministic but rather produces a *faux* violation of determinism by taking advantage of the redundancy of the surface structure theory in the sense of the many-to-one correspondence between the surface structure models and the intrinsic physical situation they describe; in particular, the models $\langle \mathcal{M}, d^*g_{ab}, d^*T_{ab}\rangle$ and $\langle \mathcal{M}, g_{ab}, T_{ab}\rangle$ in the above construction cannot threaten determinism since they are to be interpreted as different descriptions of the same physical situation. Of course, the apparatus at issue has built into it a commitment to determinism, so its application to GTR cannot be taken as part of a proof that the correct interpretation of GTR makes it a deterministic theory. The only claim being made here is that this determinism-saving move for GTR is not ad hoc but is part of a systematic approach to gauge symmetries that is taken to yield the "correct" results for pre-GTR theories.[71]

What is so clear using hindsight wisdom took Einstein many years of struggle to understand. His infamous "hole argument" can be seen as a discovery of this underdetermination problem.[72] What muddied the waters was a confusion between two senses of general covariance. *Formal general covariance* demands that the laws of motion/field equations of a theory be written in a form that makes them

[69] The plural is used here since there is a momentum constraint and a Hamiltonian constraint for every point of space.

[70] The bracket of a pair of the constraints is not always a linear combination of the constraints multiplied by a "structure constant." This failure of the constraint algebra to form a Lie algebra means that GTR is not a gauge theory in the sense of Yang-Mills. But it certainly does not mean that GTR does not contain non-trivial degrees of gauge freedom.

[71] Because these matters are surrounded by so much controversy in the philosophical literature, I want to emphasize as strongly as possible that I am not proposing a new way of looking at GTR but am simply expounding what is the standard view among general relativists; see, for example, [Wald, 1984].

[72] See [Norton, 1984] and [Stachel, 1986] for accounts of how the "hole argument" figured in Einstein's search for his gravitational field equations. And see [Rovelli, this. vol, Ch. 12] for an account of how reflecting on the lessons of the "hole argument" influenced his understanding of classical GTR and his approach to quantum gravity.

covariant under arbitrary coordinate transformations. The terminology "formal" was chosen with malice aforethought since the demand of formal general covariance is a demand on the form rather than on the content of theory. For example, Newtonian and special relativistic theories can be reformulated, without change of content, so as to meet this demand. Indeed, the fact that Newtonian and special relativistic theories can be formulated in a completely coordinate-free manner already should make it clear that coordinates cannot matter.[73] *Substantive general covariance* demands diffeomorphism invariance (e.g. that for arbitrary diffeomorphism of $\mathcal{M}$, $(\mathcal{M}, d^*g_{ab}, d^*T_{ab})$ is a model of the theory if $(\mathcal{M}, g_{ab}, T_{ab})$ is) *and* that this diffeomorphism invariance is a gauge symmetry. Again the terminology "substantive" was chosen with malice aforethought since the demand of substantive general covariance is not automatically met, without change of content, for formally generally covariant Newtonian and special relativistic theories, at least not by the lights of apparatus for treating gauge symmetries that has been touted here (see [Earman, 2006]).

What invites confusion is the fact that a spacetime coordinate transformation can be taken to indicate either a relabeling of spacetime points or as indicating a (local) diffeomorphism. In the first guise these transformations are gauge transformations of an innocent kind: they relate the various coordinate representations of the intrinsic coordinate-free objects g_{ab} and T_{ab} obtained by taking the components of these objects in different coordinate systems. But there is nothing new here as regards GTR since exactly the same story holds for intrinsic coordinate-free presentations of pre-GTR theories. In the second guise, however, these transformations may or may not be gauge transformations — it depends on the content of the theory.

When he first discovered the underdetermination problem by means of the "hole argument," Einstein took it to reveal a real and intolerable form of underdetermination. To avoid it, he thought he had to reject formal general covariance as a *desideratum* for gravitational field equations. Only after wandering in the wilderness of non-covariant equations for almost three years did he re-embrace general covariance. In effecting the re-embrace Einstein did not speak the language of gauge symmetries (the terminology and the apparatus had not been invented), so he did not say that the gauge interpretation of GTR lowers the hurdle for determinism in that it requires only the uniqueness of evolution for gauge invariant quantities. But he said what amounts to the same thing; or rather he said it for a subclass of the gauge invariant quantities of GTR — what are called "point coincidences," i.e. things like the intersection of light rays.[74]

Many philosophers have traced Einstein's path in various ways. Very few of them, however, have owned up to the implications of where the path leads. If

---

[73] In the above presentation I have intentionally used the "abstract index" notation. Thus, g_{ab} stands for a symmetric, covariant tensor field that is defined in a coordinate-free manner as a bilinear map of pairs of tangent vectors to $\mathbb{R}$. This object can be represented by its coordinate components g_{jk} in a coordinate system $\{x^i\}$. The transformations between two such representations are gauge transformations, albeit trivial ones.

[74] See [Howard, 1999] for an account of Einstein's use of this term.

determinism in GTR is saved by treating diffeomorphism invariance as a gauge symmetry, then the only "observables" (= genuine physical magnitudes) of GTR are gauge invariant quantities. This is easy enough to say, but what exactly is the nature of the gauge invariant structure that underlies the surface structure? This is a crucial issue for those physicists who pursue a quantum theory of gravity by applying some version of the canonical quantization program to GTR, for on this program it is the "observables" of classical GTR that will be turned into quantum observables in the sense of self-adjoint operators on the Hilbert space of quantum gravity. There is no standard answer to the question of how best to characterize the observables of classical GTR. But one thing is sure: none of the familiar quantities used in textbook presentations of GTR, not even scalar curvature invariants such as the Ricci scalar appearing in the field equations (17), count as observables in the sense under discussion. And more particularly, no local quantities — quantities attached to spacetime points or finite regions — are gauge invariants. In this respect the gauge-free content of the theory has a non-substantivalist flavor. Whether this content can be characterized in a way that also satisfies traditional relationalist scruples remains to be seen.

A second closely related implication of treating the diffeomorphism invariance of GTR as a gauge symmetry concerns the nature of time and change. In the Hamiltonian formalism the implication takes the form of a "frozen dynamics." Applying to the Hamiltonian constraint of GTR the doctrine that first class constraints generate gauge transformations leads directly to the conclusion that motion in GTR is pure gauge. Put another way, the instantaneous states in the Hamiltonian formulation of the theory contain redundant structure, and any two such states, where one is generated from another by solving the Hamiltonian form of EFE, are equivalent descriptions of the same intrinsic, gauge invariant situation.[75]

For those who find these implications unpalatable, the heterodox moves mentioned in Section 3.3 may be attractive. As far as I am aware, however, such heterodoxy as applied to GTR has not been seriously pursued by the physics community.

6.3 The initial value problem in GTR

For the sake of simplicity consider the initial value problem for the source-free or vacuum EFE, i.e. (17) with $T_{ab} \equiv 0$. Since these equations are second order in time, presumably the appropriate initial data consist of the values, at some given time, of the spacetime metric and its first time derivative. The technical formulation of this idea is to take an initial data set to consist of a triple (Σ, h_{ab}, k_{ab}), with the following features and intended interpretations. Σ is a three-manifold, which is to be embedded as a spacelike hypersurface of spacetime $\mathcal{M}, g_{ab}$. h_{ab} is a smooth Riemann metric on Σ, which will coincide with the metric induced on Σ by the spacetime metric g_{ab} when Σ is embedded as a spacelike hypersurface

[75]For more on the problem of time in GTR and quantum gravity, see [Belot and Earman, 1999], [Belot, this vol.], and [Rovelli, this vol.].

of $\mathcal{M}, g_{ab}$. And k_{ab} is a smooth symmetric tensor field on Σ that coincides with the normal derivative of h_{ab} when Σ is embedded as a spacelike hypersurface of $\mathcal{M}, g_{ab}$. A spacetime $\mathcal{M}, g_{ab}$ that fulfills all of these roles is said to be a *development* of the initial data set (Σ, h_{ab}, k_{ab}). If the development $\mathcal{M}, g_{ab}$ of the initial data set (Σ, h_{ab}, k_{ab}) satisfies the source-free EFE, then h_{ab} and k_{ab} cannot be specified arbitrarily but must satisfy a set of constraint equations. The existence and uniqueness result for the source-free EFE takes the following form[76]: Let (Σ, h_{ab}, k_{ab}) be an initial value set satisfying the constraint equations; then there exits a development $\mathcal{M}, g_{ab}$ of the initial data that is the unique — up to diffeomorphism — maximal Cauchy development satisfying the source-free field equations. Furthermore, g_{ab} depends continuously on the initial data (see [Hawking and Ellis, 1973] for details of the relevant topology).

Just as the proof of the well-posedness of the initial value problem for the homogeneous Maxwell equations exploits the gauge freedom in the electromagnetic potentials (see Section 4.2), so the existence and uniqueness proof for EFE exploits the idea that diffeomorphism invariance is a gauge symmetry of GTR. When the metric potentials g_{ij} (i.e. the coordinate components of the metric g_{ab}) are subjected to a gauge condition (called the harmonic coordinate condition), the EFE take the form of a system of quasi-linear, diagonal, second order hyperbolic pdes, which are known to have locally well-posed initial value problem.

That the development $\mathcal{M}, g_{ab}$ of the given initial data is a Cauchy development means that Σ is a Cauchy surface of $\mathcal{M}, g_{ab}$ (and, thus, that this spacetime is globally hyperbolic). That it is the maximal Cauchy development means that there is no proper extension of $\mathcal{M}, g_{ab}$ which is a solution of the source-free EFE and for which Σ is a Cauchy surface. The up-to-diffeomorphism qualifier to uniqueness was to be expected from the discussion of gauge freedom in the previous subsection, and in turn the presence of this qualifier shows that the heuristic discussion given there can be given precise content. Here the qualifier means that if $\mathcal{M}', g'_{ab}$ is any other maximal development satisfying the source-free EFE, then there is a diffeomorphism $d : \mathcal{M} \to \mathcal{M}'$ such that $d^* g_{ab} = g'_{ab}$.

Curie's Principle (see Section 2.3 above) would lead one to believe that a symmetry of the initial value set (Σ, h_{ab}, k_{ab}) for the vacuum EFE should be inherited by the corresponding solution. And so it is. Let $\varphi : \Sigma \to \Sigma$ be a diffeomorphism that is a symmetry of the initial data in the sense that $\varphi^* h_{ab} = h_{ab}$ and $\varphi^* k_{ab} = k_{ab}$. Then as shown by Friedrich and Rendall [2000, 216–217], if ψ is an embedding of Σ into the maximal Cauchy development determined by (Σ, h_{ab}, k_{ab}), there exists an isometry $\bar{\psi}$ of this development onto itself such that $\bar{\psi} \circ \varphi = \varphi \circ \psi$, i.e. there is an isometry of the maximal Cauchy development whose restriction to $\varphi(\Sigma)$ is ψ. Moreover, this extension of the symmetry of the initial data is unique.

The initial value problem for the sourced EFE involves not only the stress-energy tensor T_{ab} but also the equations of motion for the matter fields that give rise to T_{ab} and, in particular, the coupling of these matter fields to gravity and to each other. Whether the coupled Einstein-matter equations admit an initial value formulation

[76]This formulation is taken from Wald [1984, Theorem 10.2.2].

and, if so, whether the initial value problem is well-posed are issues that have to be studied on a case-by-case basis. For what seem to be appropriate choices of coupling, the initial value problem for the combined Einstein-Klein-Gordon equations and the Einstein-Maxwell equations have existence and uniqueness results similar to that for the source-free Einstein equations. For other cases the results are not as nice.[77]

The results mentioned above demonstrate that substantive general covariance (in the sense that diffeomorphism invariance is a gauge symmetry) is compatible with having a well-posed initial value problem. But there is clearly a tension between the two, and so one can wonder just how tightly the combination of these two requirements constrains possible laws.[78]

6.4 Cosmic censorship and chronology protection

The positive results reported in the preceding section hardly exhaust the issue of determinism in GTR. One key concern is what happens when the maximal Cauchy development $\mathcal{M}, g_{ab}$ of initial data (Σ, h_{ab}, k_{ab}) satisfying the constraint equations is not maximal simpliciter, i.e. when $\mathcal{M}, g_{ab}$ can be imbedded isometrically as a proper subset of a larger spacetime $\mathcal{M}', g'_{ab}$ satisfying the source-free EFE. The analogous issue can also be raised for the case when $T_{ab} \neq 0$. The future boundary $H^+(\Sigma)$ of the (image of) the future domain of dependence $D^+(\Sigma)$ in the larger spacetime is called the *future Cauchy horizon* of Σ; the *past Cauchy horizon* $H^-(\Sigma)$ of Σ is defined analogously.[79] Intuitively, beyond the Cauchy horizons of Σ lie the regions of spacetime where the state of things is not uniquely fixed by the given initial data on Σ; for generally if the maximal Cauchy development $\mathcal{M}, g_{ab}$ of the initial data is not maximal simpliciter, then the larger extensions for which Σ is not a Cauchy surface are not unique (even up-to-diffeomorphism).

A relatively uninteresting reason why the maximal Cauchy development might be non-maximal simpliciter is that Σ was a poor choice of initial value hypersurface. A trivial but useful example is given by choosing Σ to be the spacelike hyperboloid of Minkowski spacetime pictured in Fig. 3. Here $H^+(\Sigma)$ is the past null cone of the point p.

Some features of this example generalize; in particular, $H^+(\Sigma)$ is always a null surface generated by null geodesics. The unfortunate case can be excluded by requiring, say, that Σ be compact or that it be asymptotically flat. Of course, these conditions exclude many cases of physical relevance, but for sake of discussion let us leave them in place. Even so, the maximal Cauchy development may fail to be maximal simpliciter for more interesting and more disturbing reasons.

[77] For comprehensive reviews of what is known, see [Friedrich and Rendall, 2000] and [Rendall, 2002].

[78] An analysis of gauge symmetries different from the one advertised here is given in [Geroch, 2004]. He gives only two examples of laws that have an initial value formulation and that have diffeomorphism invariance as a gauge symmetry (in his sense).

[79] More precisely, $H^+(\Sigma) := \overline{D^+(\Sigma)} - I^-(D^+(\Sigma))$, and analogously for $H^-(\Sigma)$.

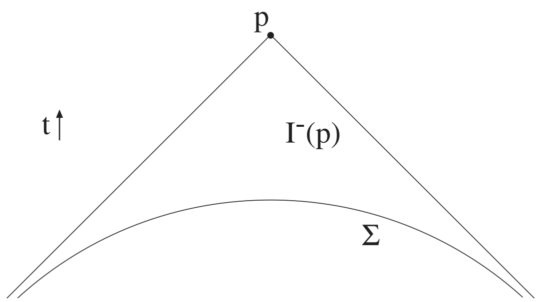


Figure 3. A poor choice of initial value hypersurface

One reason is that a spacetime can start with nice causal properties and evolve in such a way that these properties are lost. The point is illustrated by Misner's (1 + 1)-dim spacetime that captures some of the causal features of Taub-NUT spacetime, which is a solution to the source-free EFE. The Σ in Fig. 4 is a compact spacelike slice in the causally well behaved Taub portion of the spacetime, and its future Cauchy horizon $H^+(\Sigma)$ is a closed null curve. Crossing over this horizon takes one into a region of spacetime where there are closed timelike curves.

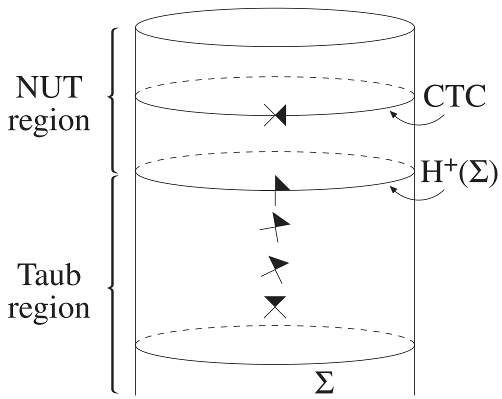


Figure 4. Misner spacetime

Another reason that the maximal Cauchy development may fail to be maximal simpliciter is illustrated in Fig. 5 which shows a non-compact asymptotically flat spacelike slice Σ on which a spherically symmetric ball of matter starts to undergo gravitational collapse. After a finite time the density of collapsing matter becomes infinite, creating a curvature singularity that is pictured as a timelike line in the figure. Strictly speaking, however, it makes no sense to call the singularity a timelike line since the singularity is not part of the spacetime.[80] But this makes no difference to the main point of relevance here; namely, a causal curve that terminates at a point to the future of $H^+(\Sigma)$ and that is extended into the past may fail to reach Σ, not because it has a past endpoint or because it gets trapped on $H^+(\Sigma)$ (as can happen in the spacetime of Fig. 4) but because it "runs into a singularity" or, better (since the singularity is not part of the spacetime), because it "runs off the edge of the spacetime."

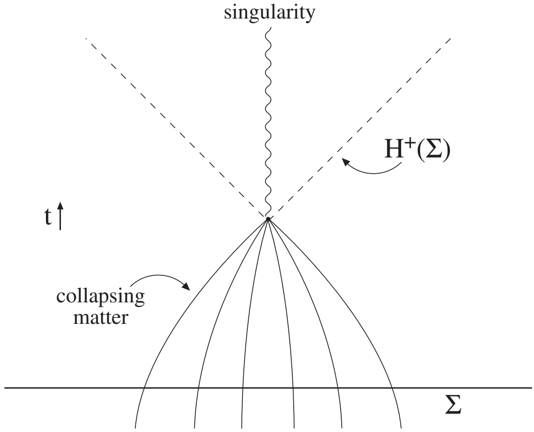


Figure 5. The development of a naked singularity in spherical gravitational collapse

It is known that EFE plus the imposition of the various energy conditions on T_{ab} discussed above do not suffice to prevent the kind of pathologies indicated by Figs. 4 and 5. But in all of such known examples there is something suspicious: either the matter fields involved are not "fundamental," i.e. even when gravity is turned off these matter fields are not well behaved in the sense that in Minkowski spacetime the initial value problem for their equations of motion do not admit global existence and uniqueness results (see Section 4.2), or else the initial conditions that eventuate in the singularity are very special, e.g. the initial configuration of matter in Fig. 5 required to produce the singularity must be perfectly spherically symmetric. One might conjecture that what holds for these examples holds generally: Consider a fundamental matter field that can serve as a

[80]One could try to attach the singular points as boundary points of the spacetime manifold. However, the extant prescriptions for doing this lead to counterintuitive features, e.g. the singular points need not be Hausdorff separated from interior points of the manifold; see [Geroch *et al.*, 1982].

source for gravitation. Then the subset of initial data for the Einstein-matter field equations for which the unique (up to diffeomorphism) maximal Cauchy development is not maximal simpliciter is of measure zero in the full space of such data, assuming poor choices of initial value hypersurfaces are excluded. To make this vague claim into a precise conjecture would require a specification of what matter fields are to be counted as fundamental, a specification of a suitable measure on the space of initial data, and a non-question begging specification of what counts as a poor choice of initial value hypersurface. The aimed-for conjecture is referred to as Penrose's *cosmic censorship conjecture.*

Less sweeping versions of the conjecture might focus specifically on one or another of the two types of pathologies illustrated in Figs. 4 and 5. Hawking's *chronology protection conjecture* aims to show that cases where closed timelike curves develop from a causally innocent past are highly non-generic among solutions to Einstein-fundamental-matter-field equations. The *weak cosmic censorship conjecture* aims to show that in generic solutions with asymptotically flat spacetimes, singularities are not "naked" in the sense of being visible to observers at infinity because whatever singularities develop (say, in gravitational collapse) are hidden inside of the event horizons of black holes which serve as one-way causal membranes that shield external observers from any of the pathologies of indeterminism that might develop within the horizon. Some progress has been made in formulating and proving precise versions of chronology protection and weak cosmic censorship, but the juries are still out on strong cosmic censorship.[81]

6.5 Predictability in general relativistic spacetimes

In Section 4.3 it was seen that the structure of Minkowski spacetimes has a double-edged quality with respect to determinism and predictability: while this structure makes possible clean examples of determinism, it also makes it impossible for embodied observers who must gather their information about initial conditions by means of causal interactions with the world to use determinism to perform genuine predictions. The point was formalized by defining the domain of predictability $P(q)$ of a point $q \in \mathcal{M}$ of a spacetime $\mathcal{M}, g_{ab}$ and noting that in Minkowski spacetime $P(q) = \varnothing$ for every q. Non-empty domains of predictability are obtained in the modified version of Minkowski spacetime with compactified space slices illustrated in Fig. 2. A feature of this case generalizes to arbitrary general relativistic spacetimes; namely, if the spacetime $\mathcal{M}, g_{ab}$ possesses a Cauchy surface Σ such that $\Sigma \subset I^{-}(q)$ for some $q \in \mathcal{M}$, then Σ is compact. Since a spacetime with a Cauchy surface Σ is diffeomorphically Σ x $\mathbb{R}$, the kind of complete predictability that comes with having $\Sigma \subset I^{-}(q)$ for some q is possible only in a spatially finite universe. The converse is not true: the existence of a compact Cauchy surface does not guarantee that there is a Cauchy surface Σ such that $\Sigma \subset I^{-}(q)$ for some q, de

[81]For an overview of progress on cosmic censorship, see [Chruściel, 1992; Isenberg, 1992; [Penrose, 1998; Wald, 1998]. And for an overview of progress on chronology protection, see [Earman *et al.*, 2005].

Sitter spacetime providing a relevant counterexample. Many interesting features of predictability in general relativistic spacetime are studied in [Hogarth, 1993].

6.6 Determinism and computability in general relativistic spacetimes

A *Plato machine* for gaining mathematical knowledge about an unresolved conjecture of number theory, whose prenex normal form is $(\forall n_1)(\forall n_2)...(\forall n_m)F(n_1, n_2, ..., n_m)$ or $(\exists n_1)(\exists n_2)...(\exists n_m)F(n_1, n_2, ..., n_m)$ with F recursive, can be conceptualized as an ordinary Turing machine run in Zeno fashion: enumerate the m-tuples of natural numbers and have the computer check in the first 1/2 minute whether F holds of the first tuple, check in the next 1/4 minute whether F holds of the second tuple, etc. At the end of the minute the truth of the conjecture is settled. Despite various claims to the contrary, I see no conceptual incoherency in such a device. But STR militates against the physical instantiation of such a device since the Zeno speed up would seem to require that some of the parts of the device must eventually move faster than the speed of light.[82]

General relativistic spacetimes seem to open the possibility of creating the functional equivalent of a Plato machine without Zeno tricks and without running afoul of the prohibition on superluminal propagation. Consider a spacetime with the following features. First, there is a timelike half-curve γ_1 with past endpoint, no future endpoint, and an infinite proper length. Second, there is another timelike half-curve γ_2 with past endpoint p and a point $q \in \gamma_2$ such that the proper time elapsed along γ_2 from p to q is finite and such that $\gamma_1 \in I^-(q)$. Such a spacetime has been dubbed a *Malament-Hogarth spacetime*. The theorems of any recursively axiomatizable theory — say, Zermelo-Frankel set theory — can be recursively enumerated, and a device whose worldline is γ_1 can utilize a Turing machine to effectively check each of these theorems to see one has the form "$0 = 1$". The device can be programmed to send out a signal — "Eureka!"— to an observer whose world line is γ_2 just in case "$0 = 1$" is found among the theorems. Assuming that the observer γ_2 is aware of this arrangement, she gains knowledge of the consistency/inconsistency of ZF: she knows that ZF is consistent just in case she has not received a "Eureka!" signal by the time she reaches the point q.

Similar arrangements can be used to "decide," at least in principle, Turing undecidable questions and to "compute" Turing uncomputable functions (see [Hogarth, 1994]). They, thus, threaten to falsify the physical Church-Turing thesis which asserts, roughly, that any physical computing device can be simulated by a Turing machine (see [Etsei and Németi, 2002] for a careful formulation of this thesis). In contrast to the original Church-Turing thesis which belongs to mathematical logic, the physical Church-Turing thesis lies in the borderland of mathematical logic and physics (see [Deutsch *et al.*, 2000]), and it is much harder to evaluate, especially if it is understood to require the physical realizability of the devices that implement the bifurcated supertask. Here I will confine myself to a few remarks

[82]Perhaps conflict with STR can be avoided by Zeno shrinking the parts, but this maneuver may run afoul of quantum restrictions.

on this matter and refer the interested reader to Németi and David [2005] for a fuller discussion.

Malament-Hogarth spacetimes are among the solutions of EFE — e.g. Reissner-Nordström spacetime and (the universal covering spacetime of) anti-De Sitter spacetimes. These particular spacetimes do not involve causal anomalies in the sense that they admit global time functions. However, all Malament-Hogarth spacetimes fail to be globally hyperbolic. Indeed, it can shown of such spacetimes that if $\Sigma \subset \mathcal{M}$ is any spacelike hypersurface such that the above defined γ_1 lies in $I^+(\Sigma)$, then any Malament-Hogarth point q whose chronological past contains γ_1 must lie on or beyond $H^+(\Sigma)$ (see Lemma 4.3 of [Earman, 1995, 117]). The possibility of non-deterministic influences, which might open the possibility that γ_1 receives a false "Eureka!" message, seems to undermine the use of Malament-Hogarth spacetimes for gaining knowledge in the sense of certainty. However, one should not draw hasty conclusions here since, as discussed in the following subsection, it is possible to have deterministic dynamics for fields propagating on a non-globally hyperbolic spacetime. Also it might seem that the problem is avoided by the fact that it can be arranged so that any signal from γ_1 arrives at γ_2 before the Malament-Hogarth point q and, thus, within $D^+(\Sigma)$. But since a "Eureka!" message can arrive arbitrarily close to q, the receiver must possess arbitrarily accurate discriminatory powers to separate signals that arrive before q from the potentially false signals that arrive after q.

6.7 The possibility of deterministic dynamics in non-globally hyperbolic spacetimes

For fields that propagate on a general relativistic spacetime, the failure of global hyperbolicity can undermine the initial value problem. For example, it is known that in generic two-dimensional spacetimes with closed timelike curves (CTCs) the scalar wave equation may fail to have smooth solutions or else may admit multiple solutions for the same initial data specified on a spacelike hypersurface. But remarkably, existence and uniqueness results have been proven for some four-dimensional spacetimes with CTCs (see [Friedman, 2004] for a review).

For spacetimes that do not have such blatant causal anomalies as CTCs but which nevertheless fail to be globally hyperbolic, Hilbert space techniques can sometimes be used to cure breakdowns in existence and uniqueness.[83] Consider a general relativistic spacetime $\mathcal{M}, g_{ab}$ that is static and possesses a global time function. The first condition means that there is a timelike Killing field V^a that is hypersurface orthogonal.[84] The second condition can be guaranteed by choosing

[83] The use of Hilbert space techniques to study problems in classical physics was pioneered by Koopman [1931]. However, Koopman's approach assumes determinism at the classical level and then shows how to represent this determinism as unitary evolution on a Hilbert space.

[84] The Killing condition is $\nabla_{(c}g_{ab)} = 0$ where ∇_a is the covariant derivative operator compatible with g_{ab}. Staticity guarantees that locally a local coordinate system (x^α, t) can be chosen so that the line element takes the form $ds^2 = g_{\alpha\beta}(x^\gamma)dx^\alpha dx^\beta - g_{44}(x^\gamma)dt^2$. Cf. Malament, this volume, section 2.7.

a spacelike hypersurface Σ orthogonal to V^a and by requiring that every integral curve of V^a meets Σ in exactly one point. Then every point $p \in \mathcal{M}$ can be labeled by (x, t), where $x \in \Sigma$ is the point where the integral curve of V^a through p meets Σ, and t is the value of the Killing parameter along this integral curve. Such a causally well behaved spacetime can nevertheless fail to be globally hyperbolic because, intuitively speaking, it possess a naked, timelike singularity. (To help fix intuitions, think of Minkowski spacetime with a timelike world tube removed. Or the reader familiar with GTR can think of the negative mass Schwarzschild solution to EFE, which is static and contains a timelike naked singularity at $r = 0$.) Now consider a massive $m \geq 0$ scalar field ϕ propagating on this background spacetime in accord with the Klein-Gordon equation (13). For the type of spacetime in question this equation can be rewritten in the form

$$(17) \quad \frac{\partial^2 \phi}{\partial t^2} = -A\phi$$

where t is the Killing parameter (see [Wald, 1980a], [Horowitz and Marolf, 1995]). The differential operator A can be considered a Hilbert space $\hat{A}$ operator acting on $L^2_{\mathbb{R}}(\Sigma, d\vartheta)$, where $d\vartheta$ is the volume element of Σ divided by $\sqrt{-V^a V_a}$. With the domain initially taken to be $C_0^\infty(\Sigma)$, $\hat{A}$ is a positive symmetric operator. The proposal is to replace the partial differential equation (18) with the ordinary differential equation

$$(18) \quad \frac{d^2 \phi}{dt^2} = -\hat{A}\phi$$

where the time derivative in (19) is a Hilbert space derivative. Since the Hilbert space operator $\hat{A}$ is real it has self-adjoint extensions, and since $\hat{A}$ is positive the positive square root of a self-adjoint extension $\hat{A}_e$ can be extracted to give

$$(19) \quad \phi(t) := \cos(\sqrt{\hat{A}_e} t)\phi(0) + \sin(\sqrt{\hat{A}_e} t)\dot{\phi}(0)$$

which is valid for all t and all $\phi(0), \dot{\phi}(0) \in \mathcal{H}$. Since $\phi(t)$ is a solution throughout the spacetime of the Klein-Gordon equation (13) and since it is the unique solution for given initial data $\phi(0), \dot{\phi}(0)$ on Σ, it provides (relative to the chosen self-adjoint extension) a deterministic prescription for the dynamics of the Klein-Gordon field. There are other possible prescriptions for obtaining the dynamics of ϕ, but Ishibashi and Wald [2003] have shown that the one just reviewed is the only one satisfying the following set of restrictions: it agrees locally with (18); it admits a suitable conserved energy; it propagates the field causally; and it obeys time translation and time reflection invariance. If the Hilbert space operator $\hat{A}$ is essentially self-adjoint, then the unique self-adjoint extension $\hat{A}_e$ provides *the* dynamics for the ϕ field satisfying the said restrictions. And this dynamics is fully deterministic despite the fact that the background spacetime on which the field propagates is not globally hyperbolic. Not surprisingly, however, $\hat{A}$ fails to be essentially self-adjoint for many examples of static but non-globally hyperbolic spacetimes, and unless further restrictions are added to single out one of the self-adjoint extensions, no unambiguous dynamics is specified by the above procedure. But remarkably, Horowitz and Marolf [1995] have provided examples of static, non-globally hyperbolic spacetimes where $\hat{A}$ *is* essentially self-adjoint, and in these cases the above prescription produces a dynamics of the ϕ field that is fully deterministic despite the presence of naked singularities.

7 DETERMINISM IN RELATIVISTIC QFT

Ordinary QM starts from a classical mechanical description of a system of particles — specifically, a Hamiltonian description — and attempts to produce a quantized version. Similarly, QFT starts from a classical relativistic description of a field and attempts to produce a quantized version. However, some classical fields do not lend themselves to a QFT that physicists find acceptable. Consider, for example, the non-linear wave equation (13) as a candidate for describing boson-boson interactions. A heuristic quantization procedure leads to the conclusion that there is no lowest energy state, leaving the system vulnerable to radiative collapse. On these grounds quantum field theorists have categorized the hypothetical interaction as "not physically realizable" (see [Baym, 1960]). That difficulties are encountered in QFT is perhaps not surprising when it is realized that the field in question is ill-behaved at the classical level in that regular initial data can pick out solutions that develop singularities within a finite amount of time. Is it plausible that deterministic behavior at the classical relativistic level can serve as a selection principle for what fields it is appropriate to quantize?

Determinism also plays a more constructive role in QFT. In ordinary QM, quantization involves the choice of a suitable representation of the canonical commutation relations $[\hat{x}_j, \hat{p}_k] = i\delta_{jk}$ (CCR). Since unbounded operators are involved, this form of the CCR only makes sense when the domains of the operators are specified. Such worries can be avoided by working with the Weyl, or exponentiated, form of the CCR, which also makes available the Stone-von Neumann theorem: for a finite number of degrees of freedom, the irreducible strongly continuous representations of the Weyl CCR are all unitarily equivalent — in fact, all are equivalent to the familiar Schrödinger representation. This theorem no longer applies when there are an infinite number of degrees of freedom, as in QFT, a feature of QFT that raises a number of interesting interpretational issues that are not relevant here. What is relevant is the fact that the construction of the CCR algebra for, say, the Klein-Gordon field in Minkowski spacetime, makes essential use of the deterministic propagation of this field (see [Wald, 1994]). This construction can be generalized to a Klein-Gordon field propagating in an arbitrary general relativistic background spacetime that is globally hyperbolic since the deterministic nature of the propagation carries over to the more general setting.

For a non-globally hyperbolic spacetime $\mathcal{M}, g_{ab}$ it is still the case that for any $p \in \mathcal{M}$ there is a neighborhood $\mathcal{N}(p)$ such that $\mathcal{N}, g_{ab}|_{\mathcal{N}}$, considered as a spacetime in its own right, is globally hyperbolic, and thus the field algebra $\mathcal{A}(\mathcal{N})$ for this mini-spacetime can be constructed by the usual means. One can then ask whether these local algebras can be fitted together to form a global algebra $\mathcal{A}(\mathcal{M})$ with the natural net properties (e.g. each such $\mathcal{A}(\mathcal{N})$ is a subalgebra of $\mathcal{A}(\mathcal{M})$, and if $\mathcal{N}_1 \subset \mathcal{N}_2$ then $\mathcal{A}(\mathcal{N}_1)$ is a subalgebra of $\mathcal{A}(\mathcal{N}_2)$). Kay (1992) calls the spacetimes for which the answer is affirmative *quantum compatible*, the idea being that non-quantum compatible spacetimes are not suitable arenas for QFT. A variety of non-globally hyperbolic spacetimes are not quantum compatible, e.g. 2-dim cylindrical spacetimes obtained from two-dimensional Minkowski spacetime by identifications along the time axis. But, remarkably, some acausal spacetimes have been shown to be quantum compatible (see [Fewster and Higuchi, 1996] and [Fewster, 1999]).

8 DETERMINISM AND QUANTUM GRAVITY

Arguably the biggest challenge in theoretical physics today is to combine the insights of GTR and QFT so as to produce a quantum theory of gravity (see [Rovelli, this vol.]). Some inkling of what this sought after theory will yield can perhaps be gained from semi-classical quantum gravity, which is a kind of shot-gun marriage of GTR and QFT. Semi-classical means that there is no attempt to quantize the metric of spacetime, but rather than merely treating a general relativistic spacetime as a fixed background on which quantum fields propagate (as in the preceding section), an attempt is made to calculate the back-reaction on the metric by inserting the quantum expectation value of the (renormalized) stress-energy in place of the classical stress-energy tensor on the right hand side of EFE (17). Although there may be no consistent theory underlying such a procedure, good theoretical physicists know how to extract usable information from it. Perhaps the most spectacular extraction is Hawking's conclusion that a black hole is not black but radiates exactly like a black body at temperature proportional to the surface gravity of the black hole. This *Hawking effect* is taken as confirmation that the formula for black hole entropy,[85] which had been derived by independent means, is more than a formal expression; it shows that black hole entropy is the ordinary thermodynamic entropy of a black hole (see [Wald, 1994]).[86] Theoretical physicists of different schools are in agreement that this is a stable result that has to be accommodated by an adequate quantum theory of gravity. But from this point on, the disagreements increase to the point of stridency.

The Hawking effect means that, when quantum effects are taken into account, black holes are not stable objects because the Hawking radiation must be accompanied by a diminution of the mass of the black hole. Presumably, as this process

[85] $S_{bh} = \frac{kc^3}{4G\hbar}A$, where A is the surface area of the black hole.

[86] The Hawking effect is related to, but distinct from, the *Unruh effect*. The latter effect is analyzed in terms of the apparatus of quantum statistical mechanics discussed in [Emch, this vol.]. In Minkowski spacetime the essence of the Unruh effect is that what an observer uniformly accelerated through the Minkowski vacuum experiences is described by a KMS state. The Unruh effect has been generalized to general relativistic spacetime; see [Kay and Wald, 1991].

goes deeper and deeper into the quantum regime, the semi-classical calculation will eventually break down. But *if* the continuation of the calculation can be trusted, then in the fullness of time the black hole will completely evaporate. (The estimated evaporation time for a black hole of solar mass is the order of 10^{67} years, much greater than the age of the universe. But this is no problem in a universe with an infinite future, as the latest cosmological measurements indicate is the case for our universe.) And *if* the result of the evaporation can be described by a classical general relativistic spacetime, the result is a momentarily naked singularity and a breakdown in global hyperbolicity, as is indicated in Fig. 6.[87] So even if some form of cosmic censorship holds for classical GTR, quantum effects seem to undo it.

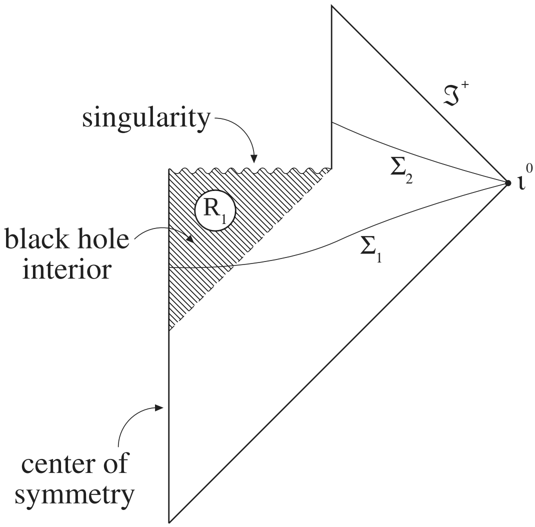


Figure 6. Conformal representation of black hole evaporation

Having gotten this far it is not difficult to establish that if at a time Σ_1 prior to the evaporation of the black hole the quantum field is in a pure state and if Hawking radiation has established correlations between relatively spacelike regions, such as the region R_1 in the black hole interior (see Fig. 6) and the region consisting of a "sandwich" about the post evaporation time Σ_2, then the state of the quantum field will be mixed at a post evaporation time Σ_2.[88] Since a pure-to-mixed state transition is necessarily non-unitary, the upshot is a loss of unitarity.[89]

[87]Following the conventions of conformal diagrams (see [Hawking and Ellis, 1973]), $\mathfrak{I}^+$ denotes future null infinity (the terminus of outgoing null rays), and ι^o denotes spatial infinity.

[88]This can be rigorously established in the algebraic formulation of QFT; see [Earman, 2002].

[89]And, incidentally, there is also a violation of time reversal invariance; see [Wald, 1986] and [Earman, 2002].

This "information loss paradox," as it is often referred to in the physics and the popular literature, has evoked an amazing variety of reactions; see [Belot *et al.*, 1999] for an overview. Most notable are the reactions from those who are so desperate to avoid the conclusion that they are willing to deploy "black hole complementarity"[90] and thereby abandon the mainstream reading of relativity theory, namely, that what the theory teaches us is that there is an intrinsic observer-independent reality — the very opposite of vulgar relativism that has it that everything is relative-to-an-observer.

But stepping back from the fray allows one to see that there is no need for such desperate measures. The pure-to-mixed evolution that is at the heart of the "paradox" need not be seen as a breakdown of quantum theory.[91] Nor is it surprising that consequences labeled 'paradoxical' flow from loss of global hyperbolicity. What needs to be questioned is whether this loss of global hyperbolicity is a plausible expectation of quantum gravity. Semi-classical quantum gravity suggests such a loss, but this way of bringing GTR and QFT together is at best a stepping stone to a full theory of quantum gravity. And just as ordinary QM showed the ability to smooth away singularities of classical mechanics, so the correct theory of quantum gravity may show the ability to smooth away the singularities of classical GTR.

Some positive indications come from the work of string theorists who are able to point to mechanisms that can smooth out singularities in classical general relativistic models; for example, Johnson *et al.* [2000] show that brane repulsion smooths out a class of naked singularities dubbed the *repulsion*. String theorists can also give a back door argument for exclusion of some types of classical singularities: postulate or prove that the sought after M-theory gives a stable ground state, and then note that this rules out analogues of the negative mass Schwarzschild solution and the like.

Other encouraging results come from loop quantum gravity (LQG), which aims to produce a quantum theory of gravity by applying to GTR a version of the canonical quantization based on a new set of canonical variables introduced by Amitaba Sen and exploited by Abay Ashtekar.[92] In the Friedmann-Robertson-Walker big bang models of classical GTR the scale factor a of spacetime goes to zero as the big bang singularity is approached, and the curvature blows up since

---

[90] Consider the case in STR of two inertial observers, O and O', who describe an ambient electromagnetic field using electric and magnetic fields $(\mathbf{E}, \mathbf{B})$ and $(\mathbf{E}', \mathbf{B}')$ respectively. There is a translation between the two descriptions which gives $\mathbf{E}'$and $\mathbf{B}'$ as functions of $\mathbf{E}$, $\mathbf{B}$, and the relative velocity of O and O' and vice versa with O and O' exchanged. The existence of such a translation follows from the fact that there an intrinsic, observer independent reality — in this case, the electromagnetic field as specified by the Maxwell tensor field. This tensor field is independent of coordinate systems, reference, frame, and observers. Contracting it with different velocity fields, representing the motions of different observers, results in different descriptions in terms of electric and magnetic fields. Whatever else it means, the "complementarity" part of "black hole complementarity" means that the different descriptions of an evaporating black hole given by two observers, one who falls through the black hole horizon and one who remains outside the horizon, are not related in the way the descriptions of O and O' are related.

[91] See [Wald, 1994, 181–182] and [Belot *et al.*, 1999].

[92] See [Rovelli, 2004] and [this vol.] for surveys of loop quantum gravity.

it scales as $1/a^2$.[93] Since there is no physically motivated way to extend such a solution through the initial singularity, the question of what happens "before" the big band belongs to theology or science fiction rather than science. The situation is startlingly different in LQG. Corresponding to the classical quantity $1/a$ there is a self-adjoint operator, acting on the Hilbert space of spatially homogeneous, isotropic quantum kinematical states, and its spectrum is bounded from above, giving a first indication that the classical singularity has been removed (see [Bojowald, 2001]). A complete proof of removal would require that the quantum dynamics gives an unambiguous evolution through the classical singularity. In LQG the "dynamics" is obtained by solving the Hamiltonian constraint equation, which restricts the physically allowed states. For the case at issue this constraint equation comes in the form of a difference equation rather than a differential equation. If the scale factor a is regarded as a "clock variable," then the constraint equation provides a "time evolution" of the quantum state through discrete steps of the clock variable. The key point is that this evolution equation does determine a unique continuation through the classical singularity.[94] However, what happens at the classical singularity is undetermined because the coefficient corresponding to this stage decouples from the other coefficients in the evolution equation (see [Ashtekar and Bojowald, 2003] for details).

From the point of view of determinism this last result means that the situation is somewhat ironic. Determinism is not threatened in classical GTR by the initial big bang singularity of the Friedmann-Robertson-Walker models because these models are globally hyperbolic, and because there is no physically motivated way to extend through the initial singularity. In LQG the initial singularity is banished both in the sense that curvature remains bounded and in the sense that there is a sensible way to extend through the classical singularity. But the price to be paid is a loss of determinism in LQG at the classical singularity, which can be seen as a Cheshire grin of the classical singularity.

Recently LQG has been used to resolve black hole singularities, leading to a new perspective on the Hawking information loss paradox in which Fig. 6 is not a valid depiction of black hole evaporation (see [Ashtekar *et al.*, 2005]). It is argued that, analogously to the FRW case, the quantum evolution continues through the classical singularity.[95] The new picture is not one in which global hyperbolicity is restored; indeed, that concept is not meaningful since what replaces the classical singularity is a region which cannot be described even approximately by the spacetime geometry of classical GTR. Nevertheless, it is argued that in the quantum

---

[93] The line element of FRW models can be written in the form $ds^2 = a(t)d\sigma^2 - dt^2$, where $d\sigma^2$ is the spatial line element.

[94] But see [Green and Unruh, 2004] where it is shown that in a spatially closed FRW model, the use of the scale factor as a "clock variable" is problematic. And the situation in inhomogeneous cosmologies is much more delicate and complicated; see [Brunnemann and Thiemann, 2006a; 2006b].

[95] As in the FRW case, the Hamiltonian constraint equation becomes a difference equation. The "quantum evolution" comes from this equation by choosing a suitable "clock variable" and then following the quantum state through discrete steps of the clock variable.

evolution a pure state remains pure and, in this sense, no information is lost. In its present form the argument has a heuristic character, and detailed calculations are needed to make it rigorous.

9 CONCLUSION

Is the world deterministic? Without the aid of metaphysical revelation, the only way we have to tackle this question is to examine the fruits of scientific theorizing. We can thus set ourselves the task of going through the theories of modern physics and asking for each: If the world is the way it would have to be in order for the theory to be true, is it deterministic? One of the things we discovered is that this task is far from straightforward, for the way in which theories are interpreted is colored by our attitudes towards determinism. For example, the unwillingness to see determinism fail at the starting gate in Newtonian gravitational theory militates in favor of taking gravitation to be a direct interparticle interaction and against assigning independent degrees of freedom to the Newtonian gravitational field. And an unwillingness to see determinism fail at the starting gate in GTR leads to the rejection of a naively realistic interpretation of the textbook version of the theory's description of spacetime and to the acceptance of diffeomorphism invariance as a gauge symmetry — which entails that none of the variables used in textbook presentations is counted as a genuine (= gauge invariant) physical magnitude.

The fortunes of determinism are too complicated to admit of a summary that is both short and accurate, but roughly speaking the story for classical (= non-quantum theories) is this. In Newtonian theories determinism is hard to achieve without the aid of various supplementary assumptions that threaten to become question-begging. For special relativistic theories determinism appears so secure that it is used as a selection criterion for "fundamental fields." GTR, under the appropriate gauge interpretation, is deterministic locally in time; but whether it is deterministic non-locally in time devolves into the unsettled issues of cosmic censorship and chronology protection.

Quantum physics is the strangest and most difficult case. Ordinary QM is in some respects more deterministic than Newtonian mechanics; for example, QM is able to cure some of the failures of Newtonian determinism which occur either because of non-uniqueness of solutions or the breakdown of solutions. But the fortunes of determinism in QM ultimately ride on unresolved interpretational issues. The main driving force behind these issues is the need to explain how QM can account for definite outcomes of experiments or more generally, the apparent definiteness of the classical world — an ironic situation since QM is the most accurate physical theory yet devised. Some of the extant responses to this explanatory challenge would bury determinism while others give it new life.

A new arena for testing the mettle of determinism is provided by the nascent quantum theories of gravity. There are some preliminary indications that just as ordinary QM was able to smooth out singularities of Newtonian mechanics, so quantum gravity effects may smooth out singularities of classical GTR. If this smoothing ability is broad enough it would alleviate worries that there are analogues in quantum gravity of breakdowns in determinism in classical GTR associated with failures of cosmic censorship. Quantum gravity will likely put a new face on the measurement problem and related interpretational issues that arise in ordinary QM. But it is too early to say whether this new face will smile on determinism.

AKNOWLEDGMENT

It is a pleasure to acknowledge the help and advice I have received from Gordon Belot, Michael Dickson, Klaas Landsman, Carlo Rovelli, and Laura Ruetsche. Special thanks are due to Jeremy Butterfield for numerous improvements. But, needless to say, there should be no presumption that any of the above mentioned share the opinions expressed herein.

BIBLIOGRAPHY

[Albert, 1992] D. Albert. *Quantum Mechanics and Experience*. Cambridge, MA: Harvard University Press, 1992.

[Albert, 2000] D. Albert. *Time and Chance*. Cambridge, MA: Harvard University Press, 2000.

[Anderson and Thompson, 1992] I. Anderson and G. Thompson. *The inverse problem of the calculus of variations for ordinary differential equations. Memoirs of the American Mathematical Society*, No. 473. Providence, RI: American Mathematical Society, 1992.

[Ashtekar and Bojowald, 2005] A. Ashtekar and M. Bojowald. Black hole evaporation: a paradigm. 2005. *gr-qc/0504029*.

[Ashtekar *et al.*, 2003] A. Ashtekar, M. Bojowald, and J. Lewandowski. Mathematical structure of loop quantum cosmology. *Advances in Theoretical and Mathematical Physics*, 7: 233–268, 2003. *gr-qc/0304074*.

[Atmanspacher and Bishop, 2002] H. Atmanspacher and R. Bishop (eds.). *Between Chance and Choice: Interdisciplinary Perspectives on Determinism*. Thorveton: Imprint Academic, 2002.

[Ax and Kochen, 1999] J. Ax and S. Kochen. Extension of quantum mechanics to individual systems. 1999. *quant-ph/9905077*.

[Bacciagaluppi and Dickson, 1998] G. Bacciagaluppi and M. Dickson. Dynamics for modal interpretations. *Foundations of Physics,* 29: 1165–1201, 1998.

[Barbour, 1974] J. B. Barbour. Relative-distance Machian theories. *Nature*, 249: 328–329, 1974.

[Barbour, 1999] J. B. Barbour. The development of Machian themes in the twentieth century. In J. Butterfield (ed.), *The Arguments of Time*, pages 83–110. Oxford: Oxford University Press, 1999.

[Barbour and Bertotti, 1977] J. B. Barbour and B. Bertotti. Gravity and inertia in a Machian framework. *Nuovo Cimento*, 38B: 1–27, 1977.

[Bargmann, 1954] V. Bargmann. On unitary ray representations of continuous groups. *Annals of Mathematics*, 59: 1–46, 1954.

[Barrett, 1999] J. A. Barrett. *The Quantum Mechanics of Minds and Worlds*. Oxford: Oxford University Press, 1999.

[Baym, 1960] G. Baym. Inconsistency of cubic boson-boson interactions. *Physical Review*, 117: 886–888, 1960.

[Beig, 2005] R. Beig. Concepts of hyperbolicity and relativistic continuum mechanics. 2005. *gr-qc/0411092*.

[Belot and Earman, 1997] G. Belot and J. Earman. Chaos out of order: quantum mechanics, the correspondence principle, and chaos. *Studies in History and Philosophy of Modern Physics*, 28B: 147–182, 1997.

[Belot and Earman, 1999] G. Belot and J. Earman. From metaphysics to physics. In J. Butterfield and C. Pagonis (eds.), *From Physics to Philosophy*, pages 166–186. Cambridge: Cambridge University Press, 1999.

[Belot *et al.*, 1999] G. Belot, J. Earman, and L. Ruetsche. The Hawking information loss paradox: The anatomy of a controversy. *British Journal for the Philosophy of Science*, 50: 189–229, 1999.

[Bojowald, 2001] M. Bojowald. Absence of a singularity in loop quantum cosmology. *Physical Review Letters*, 86: 5227–5230, 2001.

[Brading and Brown, 2003] C. Brading and H. Brown. Symmetries and Noether's theorems. In C. Brading and H. Brown (eds.), *Symmetries in Physics: Philosophical Reflections*, pages 89–109. Cambridge: Cambridge University Press, 2003.

[Brown, 1999] H. R. Brown. Aspects of objectivity in quantum mechanics. In J. Butterfield and C. Pagonis (eds.), *From Physics to Philosophy*, pages 45–70. Cambridge: Cambridge University Press, 1999.

[Brunneman and Thieman, 2006a] J. Brunneman and T. Thiemann. On (cosmological) singularity avoidance in loop quantum gravity, *Classical nd Quantum Gravity*, 23: 1395-1427, 2006.

[Brunneman and Thieman, 2006b] J. Brunneman and T. Thiemann. Unboundedness of triad-like operators in loop quantum gravity, *Classical nd Quantum Gravity*, 23: 1429-1483, 2006.

[Bub, 1997] J. Bub. *Interpreting the Quantum World*. Cambridge: Cambridge University Press, 1997.

[Butterfield, 1998] J. Butterfield. Determinism and indeterminism. *Routledge Encyclopedia of Philosophy*, 3: 33–39. New York: Routledge, 1998.

[Callender, 2005] C. Callender. Time is the simplest (and strongest) thing. Pre-print, 2005.

[Carroll, 2004] J. Carroll (ed.). *Readings on Laws of Nature*. Pittsburgh, PA: University of Pittsburgh Press, 2004.

[Cassirer, 1956] E. Cassirer. *Determinism and Indeterminism in Modern Physics*. New Haven: Yale University Press, 1956.

[Chalmers, 1970] A. F. Chalmers. Curie's principle. *British Journal for the Philosophy of Science*, 21: 133–148, 1970.

[Chruściel, 1992] P. Chruściel. On uniqueness in the large of solutions of Einstein's equations ('strong cosmic censorship'). In M. Gotay, J. E. Marsden, and V. Moncrief (eds.), *Mathematical Aspects of Classical Field Theory*, pages 235–273. Providence, RI: American Mathematical Society, 1992.

[Courant and Hilbert, 1962] R. Courant and D. Hilbert. *Methods of Mathematical Physics*, 2 vols. New York: Interscience Publishers, 1962.

[Curie, 1894] P. Curie. Sur la symmétrie des phénomènes physiques: symmétrie d'un champ électrique et d'un champ magnétique. *Journal de Physique*, 3: 393–415, 1894.

[Cushing, 1994] J. Cushing. *Quantum Mechanics: Historical Contingency and the Copenhagen Hegemony*. Chicago: University of Chicago Press, 1994.

[Cushing and Bowman, 1999] J. Cushing and G. Bowman. Bohmian mechanics and chaos. In J. Butterfield and C. Pagonis (eds.), *From Physics to Philosophy*, pages 90–107. Cambridge: Cambridge University Press, 1999.

[Deutsch *et al.*, 2000] D Deutsch, A. Ekert, and R. Lupacchini. Machines, logic, and quantum physics. *Bulletin of Symbolic Logic*, 6: 265–283, 2000.

[Dickson, 1997] W. M. Dickson. On the plurality of dynamics: transition probabilities and modal interpretations. In R. A. Healey and G. Hellman (eds.), *Quantum Measurement: Beyond Paradox. Minnesota Studies in the Philosophy of Science*, Vol. 17, pages 160–182. Minneapolis, MN: University of Minnesota Press, 1997.

[Dickson, 1998] W. M. Dickson. *Quantum Chance and Non-Locality: Probability and Non-Locality in the Interpretations of Quantum Mechanics*. Cambridge: Cambridge University Press, 1998.

[Driver, 1977] R. D. Driver. *Ordinary and Delay Differential Equations*. New York: Springer-Verlag, 1977.

[Driver, 1979] R. D. Driver. Can the future influence the present? *Physical Review D*, 19: 1098–1107, 1979.

[Earman, 1989] J. Earman. *World Enough and Space-Time: Absolute vs. Relational Theories of Space and Time*. Cambridge, MA: MIT Press/Bradford Books, 1989.

[Earman, 1995] J. Earman. *Bangs, Crunches, Whimpers, and Shrieks: Singularities and Acausalities in Relativistic Spacetimes*. Oxford: Oxford University Press, 1995.

[Earman, 2002] J. Earman. What time reversal invariance is and why it matters. *International Journal for the Philosophy of Science*, 16: 245–264, 2002.

[Earman, 2003] J. Earman. Tracking down gauge: an ode to the constrained Hamiltonian formalism. In K. Brading and E. Castellani (eds.), *Symmetries in Physics: Philosophical Reflections*, pages 140–162. Cambridge: Cambridge university Press, 2003.

[Earman, 2004a] J. Earman. Determinism: what we have learned and what we still don't know. In J. K. Campbell, M. O'Rourke, and D. Shier (eds.), *Determinism, Freedom, and Agency*, pages 21–46. Cambridge, MA: MIT Press, 2004a.

[Earman, 2004b] J. Earman. Curie's principle and spontaneous symmetry breaking. *International Studies in Philosophy of Science*, 18: 173–199, 2004b.

[Earman, 2005] J. Earman. Essential self-adjointness: Implications for determinism and the classical-quantum correspondence. Pre-print, 2005b.

[Earman, 2006] J. Earman. Two challenges to the substantive requirement of general covaraince. *Synthese*, 148: 443-468, 2006.

[Earman and Roberts, 2006] J. Earman and J. Roberts. Contact with the nomic: a challenge to the deniers of humean supervenience. *Philosophy and Phenomenological Research*, in press, 2006.

[Earman *et al.*, 2005] J. Earman, C Smeenk, and C. Wüthrich. Do the laws of phyics forbid the operation of a time machine? Pre-print, 2005.

[Earman and Ruetsche, 2005] J. Earman and L. Ruetsche. Relativistic invariance and modal interpretations. *Philosophy of Science*, forthcoming, 2005.

[Etesi and Németi,, 2002] G. Etesi and I. Németi. Turing computability and Malament-Hogarth spacetimes. *International Journal of Theoretical Physics*, 41: 342–370, 2002.

[Fefferman, 2001] C. Fefferman. Existence and smoothness of the Navier-Stokes equation. Clay Mathematics Institute, Millennium Problems, 2001. *http://www.claymath.org/*

[Fewster, 1999] C. J. Fewster. Bisolutions to the Klein-Gordon equation and quantum field theory on 2-dimensional cylinder spacetimes. *Classical and Quantum Gravity*, 16: 769–788, 1999.

[Fewster and Higuchi, 1996] C. J. Fewster and A. Higuchi. Quantum field theory on certain non-globally hyperbolic spacetimes. *Classical and Quantum Gravity*, 13: 51–62, 1996.

[Fine, 1982a] A. Fine. Joint distributions, quantum correlations, and commuting observables. *Journal of Mathematical Physics*, 23: 1306–1310, 1982.

[Fine, 1982b] A. Fine. Hidden variables, joint probability, and the Bell inequalities. *Physical Review Letters*, 48: 291–295, 1982.

[Foias *et al.*, 2001] C. Foias, O. Manley, R. Rosa, and R. Teman. *Navier-Stokes Equations and Turbulence*. Cambridge: Cambridge University Press, 2001.

[Friedrich and Rendall, 2000] H. Friedrich and A. Rendall. The Cauchy problem for the Einstein equations. In B. G. Schmidt (ed.), *Einstein's Field Equations and their Physical Implications: Selected Essays in Honor of Jürgen Ehlers*, pages 127–223. New York: Springer-Verlag, 2000.

[Friedman, 2004] J. L. Friedman. The Cauchy problem on spacetimes that are not globally hyperbolic. In P. T. Chruściel and H. Friedrich (eds.), *The Einstein Equations and the Large Scale Behavior of Gravitational Fields*, pages 331–346. Basel: Birkhäuser, 2004.

[Geroch, 1977] R. P. Geroch. Prediction in general relativity. In J. Earman, C. Glymour, and J. Stachel (eds.), *Foundations of Spacetime Theories. Minnesota Studies in the Philosophy of Science*, Vol. 8, pages 81–93. Minneapolis, MN: University of Minnesota Press, 1977.

[Geroch, 1995] R. P. Geroch. Relativistic theories of dissipative fluids. *Journal of Mathematical Physics*, 36: 4226–4241, 1995.

[Geroch, 2004] R. P. Geroch. Gauge, diffeomorphisms, initial-value formulation, etc. In P. T. Chruściel and H. Friedrich (eds.), *The Einstein Equations and the Large Scale Behavior of Gravitational Fields*, pages 441–477. Basel: Birkhäuser, 2004.

[Geroch *et al.*, 1982] R. P. Geroch, C. B. Liang, and R. M. Wald. Singular boundaries of spacetimes. *Journal of Mathematical Physics*, 23: 432–435, 1982.

[Ghirardi *et al.*, 1986] G. C. Ghirardi, A. Rimini, and T. Weber. Unified dynamics for microscopic and macroscopic systems. *Physical Review D*, 34: 470–491, 1986.

[Green and Unruh, 2004] D. Green and W. H. Unruh. Difficulties with recollapsing models in closed isotropic loop quantum cosmology. *Physical Review D*, 103502-1-7, 2004.

[Hadamard, 1923] J. Hadamard. *Lectures on Cauchy's Problem in Linear Partial Differential Equations*, 1923. Reprinted by Dover, New York, 1952.

[Hawking and Ellis, 1973] S. W. Hawking and G. F. R. Ellis. *The Large Scale Structure of Space-Time.* Cambridge: Cambridge University Press, 1973.

[Henneaux and Teitelboim, 1992] M. Henneaux and C. Teitelboim. *Quantization of Gauge Systems.* Princeton, NJ: Princeton University Press, 1992.

[Hogarth, 1993] M. Hogarth. Predicting the future in general relativistic spacetimes. *Studies in History and Philosophy of Science*, 24B: 721–739, 1993.

[Hoefer, 2003] C. Hoefer. Causal Determinism. *Stanford Encyclopedia of Philosophy. http://plato.stanford.edu/archives/sum2005/entries/determinism-causal/*

[Hogarth, 1994] M. Hogarth. Non-Turing computers and non-Turing computability. In D. Hull, M. Forbes, and R. M. Burian (eds.), *PSA 1994*, Vol. 1, pages 126–138, 1994.

[Holland, 1993] P. R. Holland. *The Quantum Theory of Motion.* Cambridge: Cambridge University Press, 1993.

[Horowitz and Marolf, 1995] G. Horowitz and D. Marolf. Quantum probes of spacetimes singularities. *Physical Review D*, 52: 5670-5675, 1995.

[Howard, 1999] D. Howard. Point coincidences and pointer coincidences: Einstein and the invariant content of space-time theories. In H. Goenner, J. Renn, J. Ritter, and T. Sauer (eds.), *The Expanding Worlds of General Relativity, Einstein Studies*, Vol. 7, pages 463–500. Boston: Birkhäuser, 1999.

[Isenberg, 1992] J. Isenberg. Progress on strong cosmic censorship. In M. Gotay, J. E. Marsden, and V. Moncrief (eds.), *Mathematical Aspects of Classical Field Theory*, pages 403–418. Providence, RI: American Mathematical Society, 1992.

[Isham and Kuchař, 1986a] C. J. Isham and K. Kuchař. Representations of spacetime diffeomorphisms. I. Canonical parametrized field theories. *Annals of Physics*, 164: 316–333, 1986.

[Isham and Kuchař, 1986b] C. J. Isham and K. Kuchař. Representations of spacetime diffeomorphisms. II. Canonical geometrodynmamics. *Annals of Physics*, 164: 288–315, 1986.

[Ismael, 1997] J. Ismael. Curie's principle. *Synthese*, 110: 167–190, 1997.

[Ishibashi and Wald, 2003] A. Ishibashi and R. M. Wald. Dynamics in non-globally hyperbolic static spacetimes: II. General analysis of prescriptions for dynamics. *Classical and Quantum Gravity*, 20: 3815–3826, 2003.

[John, 1982] F. John. *Partial Differential Equations*, 4th ed. New York: Springer-Verlag, 1982.

[Johnson *et al.*, 2000] C. V. Johnson, A. W. Peet, and J. Polchinski. Gauge theory and the excision of repulsion singularities. *Physical Review D*, 61: 08600-1-11, 2000.

[Kato, 1995] T. Kato. *Perturbation Theory for Linear Operators.* Berlin: Springer-Verlag, 1995.

[Kay, 1992] B. S. Kay. The principle of locality and quantum field theory on (non-globally hyperbolic) curved spacetimes. *Reviews of Mathematical Physics*, Special Issue, 167–195, 1992.

[Kay and Wald, 1991] B. Kay and R. M. Wald. Theorems on the uniqueness and thermal properties of stationary, quasifree states on spacetimes with a bifurcate Killing horizon. *Physics Reports*, 207: 49–136, 1991.

[Koopman, 1931] B. O. Koopman. Hamiltonian systems and transformations in Hilbert space. *Proceedings of the National Academy of Sciences*, 17: 315–318, 1931.

[Kronz, 1998] F. M. Kronz. Nonseparability and quantum chaos. *Philosophy of Science*, 65: 5–75, 1998.

[Keller, 1957] J. B. Keller. On solutions of nonlinear wave equations. *Communications on Pure and Applied Mathematics*, 10: 523–530, 1957.

[Lanford, 1974] O. E. Lanford. Time evolution of large classical systems. In J. Moser (ed.), *Dynamical Systems, Theory and Applications. Lecture Notes in Physics*, 38: 1–207. New York: Springer-Verlag, 1974.

[Laplace, 1820] P. S. Laplace. *Théorie analytique des probabilitiés.* Paris: V. Courcier, 1820.

[Lewis, 1973] D. Lewis. *Counterfactuals.* Cambridge, MA: Harvard University Press, 1973.

[Lichtenberg and Lieberman, 1991] A. J. Lichtenberg and M. A. Lieberman. *Regular and Chaotic Dynamics*, 2nd ed. New York: Springer-Verlag, 1991.

[Malament, 1984] D. Malament. Time travel in the Gödel universe. P. D. Asquith and P. Kitcher (eds.), *PSA 1984*, Vol. 2: 91–100, 1984.

[Malament, 2004] D. Malament. On the time reversal invariance of classical electromagnetic theory. *Studies in History and Philosophy of Modern Physics*, 35B: 295–315, 2004.

[McCall, 1995] S. McCall. Time flow, non-locality, and measurement in quantum mechanics. In S. Savitt (ed.), *Time's Arrow Today*, pages 155–172. Cambridge: Cambridge University Press, 1995.

[Moore, 1990] C. Moore. Unpredictability and undecidability in dynamical systems. *Physical Review Letters*, 64: 2354-2357, 1990.

[Moore, 1991] C. Moore. Generalized shifts: Unpredictability and undecidability in dynamical systems. *Nonlinearity*, 4: 199–230, 1991.

[Nagel, 1961] E. Nagel. *The Structure of Science.* New York: Harcourt, Brace, and World, 1961.

[Németi and David, 2005] I. Németi and D. David. Relativistic computers and the Turing barrier. *Journal of Applied. Mathematics and Computation*, forthcoming, 2005.

[Norton, 1984] J. D. Norton. How Einstein found his field equations: 1912-1915. In D. Howard and J. Stachel (eds.), *Einstein and the History of General Relativity. Einstein Studies*, Vol. 1: 101–159. Boston: Birkhauser, 1984.

[Norton, 1999] J. D. Norton. A quantum mechanical supertask. *Foundations of Physics*, 29: 1265–1302, 1999.

[Olver, 1993] P. J. Olver. *Applications of Lie Groups to Differential Equations*, 2nd Ed.. New York: Springer-Verlag, 1993.

[Payne, 1975] L. E. Payne. *Improperly Posed Problems in Partial Differential Equations.* Philadelphia, PA: SIAM, 1975.

[Pearle, 1976] P. Pearle. Reduction of the state vector by a nonlinear Schrödinger equation. *Physical Review D*, 13: 857–868, 1976.

[Pearle, 1989] P. Pearle. Combining stochastic dynamical state-vector reduction with spontaneous localization. *Physical Review A*, 39: 2277–2289, 1989.

[Penrose, 1998] R. Penrose. The question of cosmic censorship. In R. Wald (ed.), *Black Holes and Relativistic Stars*, pages 103–122. Chicago, IL: University of Chicago Press, 1998.

[Pérez Laraudogoitia, 2001] J. Pérez Laraudogoitia. Supertasks. *The Stanford Encyclopedia of Philosophy*, 2001. *http://plato.stanford.edu/archives/win2001/entries/spacetime-supertasks/*

[Pérez Laraudogoitia, 2005] J. Pérez Laraudogoitia. Avoiding infinite masses. Preprint, 2005.

[Pitowsky, 1996] I. Pitowsky. Laplace's demon consults an oracle: computational complexity of prediction. *Studies in History and Philosophy of Modern Physics*, 27: 161–180, 1996.

[Planck, 1932] M. Planck. *Der Kausalbegriff in der Physik.* Leipzig: Johann Ambrosius Barth, 1932.

[Poincaré, 1952] H. Poincaré. *Science and Method.* New York: Dover Publications, 1952.

[Popper, 1982] K. R. Popper. *The Open Universe.* Totowa, NJ: Roman and Littlefield, 1982.

[Pour-el and Richards, 1981] M. B. Pour-el and J. I. Richards. The wave equation with computable initial data such that its unique solution is not computable. *Advances in Mathematics*, 39: 215–239, 1981.

[Pour-el and Richards, 1989] M. B. Pour-el and J. I. Richards. *Computability in analysis and physics.* New York: Springer-Verlag, 1989.

[Prince, 2000] G. E. Prince. The inverse problem in the calculus of variations and its ramifications. In P. J. Vassiliou and I. G. Lisle (eds.), *Geometric Approaches of Differential Equations*, pages 171–200. Cambridge: Cambridge University Press, 2000.

[Radicati, 1987] L. A. Radicati. Remarks on the early developments of the notion of symmetry breaking. In M. Doncel, A. Hermann, L. Michel, and A. Pais (eds), *Symmetries in physics (1600-1980)*, pages 195–207. Barcelona: Bellatera, 1987.

[Raju, 1994] C. J. Raju. *Time: Towards a Consistent Theory.* Dordrecht: Kluwer Academic, 1994.

[Redhead, 1987] M. Redhead. *Incompleteness, Nonlocality, and Realism.* Oxford: Clarendon Press, 1987.

[Reed and Simon, 1975] M. Reed and B. Simon. *Methods of Modern Mathematical Physics. II: Fourier Analysis, Self-Adjointness.* New York: Academic Press, 1975.

[Reichl, 1992] L. E. Reichl. *The Transition to Chaos in Conservative Classical Systems: Quantum Manifestations.* New York: Springer-Verlag, 1992.

[Reichl, 2004] L. E. Reichl. *The Transition to Chaos in Conservative Classical Systems: Quantum Manifestations*, 2nd ed. New York: Springer-Verlag, 2004,

[Rendall, 2002] A. Rendall. Theorems on existence and global dynamics for the Einstein equations. *Living Reviews in Relativity*, 2002.
http://www.livingreviews.org/Articles/Volume5/2002-6rendall

[Rice, 1953] H. G. Rice. Classes of recursively enumerable sets and their decision problems. *Transactions of the American Mathematical Society*, 74: 358–366, 1953.

[Rovelli, 2004] C. Rovelli. *Quantum Gravity.* Cambridge: Cambridge University Press, 2004.

[Saari and Xia, 1995] D. Saari and J. Xia. Off to infinity in finite time. *Notices of the American Mathematical Society*, 42: 538–546, 1995.

[Scheffer, 1993] V. Scheffer. An inivisid flow with compact support in spacetime. *Journal of Geometry and Analysis*, 3: 341-401, 1993.

[Schnirelman, 1997] A. Schnirelman. On the nonuniquenes of weak solution of the Euler equation. *Communicatins on Pure and Applied Mathematics*, 50: 1261-1268, 1997.

[Skow, 2005] B. Skow. What makes time different from space. *Noûs*, forthcoming, 2005.

[Stachel, 1986] J. Stachel. Einstein's search for general covariance: 1912-1915. In D. Howard and J. Stachel (eds.), *Einstein and the History of General Relativity. Einstein Studies*, Vol. 1: 63–100. Boston: Birkhauser, 1986.

[Stöltzner, 2003] M. Stöltzner. Vienna indeterminism: from Exner to Frank and von Mises. In P. Parrini, W. Salmon, and M. Salmon (eds.), *Logical empiricism: Historical and contemporary perspectives*, pages 194–229. Pittsburgh, PA: University of Pittsburgh Press, 2003.

[Suppes, 1991] P. Suppes. Determinism or instability. Does it matter? In G. G. Brittan (ed.), *Causality, Method, and Modalities*, pages 5–22. Dordrecht: Kluwer Academic, 1991.

[Suppes, 1993] P. Suppes. The transcendental character of determinism. In P. A. French, T. E. Uehling, and H. K. Wettstein (eds.), *Midwest Studies in Philosophy*, 18:, 242–257, 1993.

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

[Wald, 1980a] R. M. Wald. Dynamics in non-globally hyperbolic, static spacetimes. *Journal of Mathematical Physics*, 21: 2802–2805, 1980a.

[Wald, 1980b] R. M. Wald. Quantum gravity and time reversibility. *Physical Review D*, 21: 2742–2755, 1980b.

[Wald, 1984] R. M. Wald. *General Relativity.* Chicago, IL: University of Chicago Press, 1984.

[Wald, 1994] R. M. Wald. *Quantum Field Theory in Curved Spacetime and Black Hole Thermodynamics.* Chicago, IL: University of Chicago Press, 1994.

[Wald, 1998] R. M. Wald. Gravitational collapse and cosmic censorship. In C. V. Vishveshwara, B. R. Iyer, and B. Bhawal (eds.), *Black Holes, Gravitational Radiation, and the Universe.* Dordrecht: Kluwer Academic Publishers, 1998.

[Weyl, 1932] H. Weyl. *The Open World.* New Haven, CN: Yale University Press, 1932.

[Widder, 1975] D. V. Widder. *The Heat Equation.* New York: Academic Press, 1975.

[Xia, 1992] X. Xia. Existence of noncollision singularities in the N-body problem. *Annals of Mathematics*, 135: 411–468, 1992.

[Zeipel, 1908] H. von Zeipel. Sur les singularités du problème des n corps. *Arkiv för Matimatik, Astronomie och Fysik*, 4: 1–4, 1908.

[Zurek, 1998] W. H. Zurek. Decoherence, chaos, quantum-classical correspondence, and the algorithmic arrow of time. *Physica Scripta*, T76: 186–198, 1998.

[Zurek, 2003] W. H. Zurek. Decoherence, einselection, and the quantum origins of the classical. *Reviews of Modern Physics*, 75: 715–775, 2003.

INDEX