PHYSICAL THEORY AND ITS INTERPRETATION

THE WESTERN ONTARIO SERIES IN PHILOSOPHY OF SCIENCE

A SERIES OF BOOKS
IN PHILOSOPHY OF SCIENCE, METHODOLOGY, EPISTEMOLOGY,
LOGIC, HISTORY OF SCIENCE, AND RELATED FIELDS

VOLUME 72

PHYSICAL THEORY AND ITS INTERPRETATION ESSAYS IN HONOR OF JEFFREY BUB

Edited by

WILLIAM DEMOPOULOS
University of Western Ontario, Ontario, Canada
and
University of California, Irvine, U.S.A.

and

ITAMAR PITOWSKY
Hebrew University, Jerusalem, Israel

Springer

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

ISBN-10 1-4020-4875-0 (HB)
ISBN-13 978-1-4020-4875-3 (HB)
ISBN-10 1-4020-4876-9 (e-book)
ISBN-13 978-1-4020-4876-0 (e-book)

Published by Springer,
P.O. Box 17, 3300 AA Dordrecht, The Netherlands.

www.springer.com

Printed on acid-free paper

CONTENTS

WILLIAM DEMOPOULOS

PREFACE

Jeff and I met when I was a graduate student at the University of Minnesota and he was a post doctoral fellow, first in the Chemistry Department, and then in the Center for Philosophy of Science. Later we were colleagues at Western Ontario. Our friendship and collaboration owe a great deal to both these institutions.

In the mid-1960s the Center enjoyed great success under Feigl's directorship. The history of the Center has been only very partially documented. Feyerabend's recollections, reported in his *Autobiography*, and some years earlier in his remarks for Feigl's Festschrift, possess an immediacy that makes them particularly noteworthy, even if all too brief. The Center was the first American institution of its kind and a bastion of positivist and neo-positivist thought. At the time Jeff and I were there, the staff included, in addition to Feigl and Maxwell, Paul Meehl, Roger Steuwer and Keith Gunderson. There were many enthusiastic graduate students, and there was participation, on occasion, from the members of the Philosophy Department, as well as the departments of physics, psychology, mathematics and chemistry. The extent to which this (to us ideal) environment was held together by the force of Feigl's personality became evident only many years later.

The political liberalism of the Viennese Positivists was very much reflected in the philosophical atmosphere Feigl created, an atmosphere that was marked by openness, collegiality and intellectual freedom. Combined with its excellent permanent faculty and steady stream of distinguished visitors, the Center was especially well-suited to Jeff's and my early friendship, our analytic and speculative interests, and our early collaboration. This collaboration was continued when we were members of the Philosophy Department at Western Ontario.

Jeff arrived at Western first and I followed two years later, initially as his one-year replacement. History and philosophy of science in Canada received a decisive impetus from the efforts of Bob Butts when, in the late 1960s, he assumed the chairmanship of Western's Philosophy Department. Western's graduate program in philosophy of science was distinguished by a succession of highly gifted students and a sense of cooperative purpose and achievement that—especially during its first 15 years—was truly exceptional.

Philosophy of science at Western derived much of its inspiration from the emerging community of philosophers, mathematicians and physicists working in the foundations of physics. This largely mathematical orientation toward the discipline represented a divergence from the general epistemological goals of the positivist and neo-positivist movements which characterized the philosophical context and orientation of the Minnesota Center. Under Jeff's tutelage, I acquired this orientation;

it was an exhilarating change, which I avidly embraced. However I have since come to recognize the necessity of following the long path back to a considered appreciation of the earlier, positivist tradition, both for itself, and for its value in orienting foundational research.

Anyone who knows Jeff knows as well that differences in professional status are invisible to him, that he is unstinting in the time and energy he will give his friends, and that he is untiring in his determination to understand the nature of the reality quantum mechanics seeks to describe. Both Itamar and I owe a large debt to Jeff as a teacher, friend and colleague. It is gratifying to have the opportunity to repay his long-standing personal and intellectual friendship with the presentation of this collection of essays.

For this Festschrift we solicited papers from philosophers and scientists we knew Jeff to have especially profited from as interlocutors in the course of his research. We are sure to have overlooked potential contributors, and to them we apologize. By way of explanation, let me say that we wished the presentation of the news of the volume to coincide with Jeff's 63rd birthday and we wished it to be a surprise. The natural remedy for preventing errors of omission would have been to have consulted with Jeff himself or to have consulted much more widely than we did. Both of these strategies ran the risk of interfering with our desire that the volume, when announced, should be unanticipated. The initial preparation of the volume, putting it into definite form and securing abstracts, papers and promises of papers took the better part of a year. Our authors were remarkable in their discretion; we thank them whole-heartedly both for their contributions and for their personal commitment to the success of this token of appreciation of Jeff's scientific and philosophical legacy.

ACKNOWLEDGEMENT

The editors wish to thank the editor of *Contemporary Physics* Sir Peter Knight and its publisher Taylor and Francis for generously granting permission to reprint Chris Isham's paper, "Is it true; or is it false; or somewhere in between?" The paper originally appeared in *Contemporary Physics* **46**(3) (2005) 207–219. We note that the Journal's website is: http://www.tandf.co.uk/journals/titles/00107514.asp.

JOSEPH BERKOVITZ* AND MEIR HEMMO†

1. A NEW MODAL INTERPRETATION OF QUANTUM MECHANICS IN TERMS OF RELATIONAL PROPERTIES

ABSTRACT

In this paper we propose a new modal interpretation of quantum mechanics, wherein quantum states assign to systems *relational* rather than intrinsic properties. We argue that this relational modal interpretation overcomes the major problems encountered by current modal interpretations. In particular, we explain how this new interpretation addresses the measurement problem and accounts for our experience of the classical-like behavior of macroscopic systems. We further provide an outline for the dynamics of relational properties, and demonstrate how this dynamics circumvents all the no-go theorems for relativistic modal interpretations. Finally, we discuss the difficulties and the prospects of providing a genuinely relativistic modal interpretation.

1 INTRODUCTION

Two of the most central problems at the foundations of quantum mechanics are the so-called measurement problem and the question of reconciling quantum mechanics with relativity. The measurement problem arises in orthodox no-collapse quantum mechanics from the conjunction of its two principal postulates: The unitary and linear dynamics of quantum states – the Schrödinger equation in the non-relativistic case; and the so-called 'eigenstate-eigenvalue link,' according to which a system has a property corresponding to a definite value of an observable just in case its quantum state is an eigenstate of that observable. Although these postulates are very successful in accounting for the behavior of microscopic systems, they yield anomalous predictions, which are inconsistent with our experience of the classical-like behavior of macroscopic systems. For example, the Schrödinger equation entails that at the end of a z-spin measurement on a particle in a superposition of z-spin 'up' and z-spin 'down,' a macroscopic pointer will be in a superposition of distinguishable states of pointing to 'up' and pointing to 'down.' Thus, according to the eigenstate-eigenvalue link, the pointer will display no definite outcome.

The question of the compatibility of quantum mechanics with special relativity arises in different ways in collapse and no-collapse interpretations of quantum mechanics. In collapse interpretations, where collapses of quantum states are real

* Department of Philosophy, University of Maryland Baltimore County, 1000 Hilltop Circle, Baltimore, MD 21250, USA; Email: jberkov@umbc.edu
† Department of Philosophy, University of Haifa, Haifa 31905, Israel; Email: meir@research.haifa.ac.il

W. Demopoulos and I. Pitowsky (eds.), Physical Theory and its Interpretation, 1–28.

physical processes, it has been very difficult to make the collapse dynamics Lorentz covariant without singling out a preferred inertial reference frame. In no-collapse interpretations, like Bohm's theory and modal interpretations, this problem does not arise because the unitary dynamics of quantum states of isolated systems is Lorentz covariant. But such interpretations postulate the existence of additional definite properties (the so-called 'hidden-variables'), and the problem is to supply these additional properties with a genuinely relativistic dynamics.

Modal interpretations of quantum mechanics are no-collapse (typically) indeterministic interpretations that were design to solve the measurement problem and reconcile quantum mechanics with special relativity. The dynamics of quantum states of isolated systems is assumed to obey linear and unitary equations of motion, and accordingly quantum states never 'collapse.' But, orthodox no-collapse quantum mechanics is supplemented with rules for assigning additional properties, so that systems in superposition states may sometimes possess properties that correspond to one of the superposed properties. The idea is that the set of the additional properties will be rich enough to account for the occurrence of definite macroscopic events, including measurement outcomes, but sufficiently restricted so as to avoid the no-hidden-variables theorems; and the dynamics of these properties will reproduce the familiar classical-like behavior of macroscopic systems.

However, the mainstream modal interpretations fail to reproduce the classical-like behavior of macroscopic systems (Section 4). Moreover, as no-go theorems by Dickson and Clifton (1998), Arntzenius (1998) and Myrvold (2002) demonstrate, all the current modal interpretations are not genuinely relativistic. Our main aim in this paper is to consider the prospects of a relativistic modal interpretation that solves the measurement problem. In the course of our consideration, we review the measurement problem in orthodox no-collapse quantum mechanics (Section 2). We then sketch the basic interpretational rules of the mainstream versions of the modal interpretation (Section 3), and consider the measurement problem in the context of these interpretations (Section 4). Next, we introduce a new modal interpretation where the property assignment is of relational rather than intrinsic properties (Section 5), and explain how this interpretation addresses the measurement problem and more generally the problem of recovering our experience of the classical-like behavior of macroscopic systems in non-relativistic framework (Section 6). We then modify the dynamics of properties (Section 7) and show how due to its highly non-local nature the modified dynamics circumvents the no-go theorems for relativistic modal interpretations (Section 8). Further, we explain why this nonlocal character is unobservable (Section 9). Next, we consider the prospects of a relativistic relational modal interpretation (Section 10). In particular, we explore two strategies: One is to interpret the nature of the properties assigned by modal interpretations not only as related to other systems but also to spacelike hypersurfaces (Section 10.2). The other strategy is to interpret quantum states as a source of information about systems' properties in certain circumstances, which obtain when the degree of entanglement between the relevant systems is approximately zero (Section 10.3). We conclude by briefly discussing an anticipated objection to the relational modal interpretation (Section 11).

Although we focus on modal interpretations, our discussion is also relevant to other no-collapse interpretations of quantum mechanics. For the main idea of the no-go theorems for relativistic modal interpretations is similarly applicable to other no-collapse interpretations that satisfy very natural assumptions about the physical realm.

2 THE MEASUREMENT PROBLEM

In its basic form, the measurement problem may be presented in the following scheme of generic (impulsive) ideal measurement of a spin one-half observable. Let S be a measured system, M a measuring apparatus and O an observer, initially in the state:

$$|\Psi_0\rangle = (\lambda_1|\varphi_1\rangle_S + \lambda_2|\varphi_2\rangle_S)|\psi_0\rangle_M|\Phi_0\rangle_O, \tag{1}$$

where $\lambda_1, \lambda_2 \neq 0$ and $|\lambda_1|^2 + |\lambda_2|^2 = 1$, $|\varphi_i\rangle_S$ are the eigenstates of the z-spin of S, $|\psi_0\rangle_M$ is some ready state of M and $|\Phi_0\rangle_O$ is some suitable state of O's brain (and possibly sensory mechanisms) ready to read the measurement outcome. According to orthodox no-collapse quantum mechanics, during the interactions between S and M and between M and O the quantum state $|\Psi_0\rangle$ obeys the unitary and linear Schrödinger dynamics. In ideal measurements, this dynamics perfectly correlates the eigenstates of the measured observable with the eigenstates of the apparatus pointer. That is, the state (1) evolves into the state

$$|\Psi_1\rangle = (\lambda_1|\varphi_1\rangle_S|\psi_1\rangle_M + \lambda_2|\varphi_2\rangle_S|\psi_2\rangle_M)|\Phi_0\rangle_O, \tag{2}$$

where the z-spin eigenstates $|\varphi_1\rangle_S$ and $|\varphi_2\rangle_S$ are perfectly correlated with the eigenstates of the pointer observable $|\psi_1\rangle_M$ and $|\psi_2\rangle_M$. As a result, S and M get *entangled*, and accordingly their state become non-separable. Similarly, in the interaction between M and O, O becomes entangled with M:

$$|\Psi_2\rangle = \lambda_1|\varphi_1\rangle_S|\psi_1\rangle_M|\Phi_1\rangle_O + \lambda_2|\varphi_2\rangle_S|\psi_2\rangle_M|\Phi_2\rangle_O, \tag{3}$$

where $|\Phi_1\rangle_O$ and $|\Phi_2\rangle_O$ are the eigenstates of a brain observable O, associated with a perception of z-spin measurement outcome: $|\Phi_1\rangle_O$ and $|\Phi_2\rangle_O$ are associated with O's perception of pointer pointing to 'up' and pointer pointing to 'down,' respectively.[1] The eigenstate-eigenvalue link implies that the pointer observable of M and the corresponding brain observable of O have no definite values. And since the entanglement in (2) and (3) is basis-independent (these states cannot be rewritten as product states by a change of basis), the reduced states of M and O are 'improper mixtures': On pain of inconsistency, they cannot be (straightforwardly) interpreted as classical mixtures (i.e. states that assign ignorance probabilities to the various possible possessed properties). Thus, according to orthodox no-collapse quantum mechanics O fails to have a perception of a definite measurement outcome, in contradiction to our experience.[2]

It is important to bear in mind that decoherence theory (either of open systems, as in the theory of environmental decoherence[3], or of open and closed systems as in the decohering histories approach[4]) does not by itself solve the measurement problem.[5] Since the dynamics of the quantum state of isolated systems (including the environment) is linear, measurement interactions map initial product states of the form (1) into final entangled states of the form (3), no matter whether decoherence conditions are satisfied by M or O. In particular, even if the orthogonal environment states are coupled to the pointer states $|\psi_i\rangle_M$, the *reduced* state of M obtained by partial tracing cannot simply be given an ignorance interpretation; and similarly, mutatis mutandis, for O's reduced state. Thus, it follows from the eigenstate-eigenvalue link that M does not point to any definite pointer outcome and O does not possess any property that is associated with the perception of such outcome.

There are three main strategies for addressing the measurement problem. One strategy, applied in Everett-like interpretations is to bite the bullet and maintain that at the end of measurements, pointers of measurement apparatuses display definite outcomes, and are perceived as such, only relative to the so-called Everett 'branches,' where the latter are associated with the components (with non-zero amplitudes) of the Quantum superposition when written in some or other preferred way. A second strategy is to replace the unitary and linear dynamics of quantum states by a suitable collapse dynamics, as in the dynamical models for spontaneous localization (see, for example, Ghirardi, Rimini and Weber 1986 and Ghirardi, Pearle and Rimini 1990). A third strategy, applied in hidden-variables interpretations (such as Bohm's (1952) theory), is to leave intact the unitary and linear dynamics and modify the property assignment of the orthodox theory, so that at the end of measurements pointers will display definite outcomes in the entangled states (2) and (3), and the relevant brain observables have definite values that are associated with states of mind of perceiving definite outcomes in the entangled state (3). Modal interpretations belong to this latter approach.

3 MODAL INTERPRETATIONS: AN OVERVIEW

3.1 The property assignment

Modal interpretations of quantum mechanics and orthodox collapse quantum mechanics differ in three important respects. First, modal interpretations are no-collapse interpretations: They postulate that the dynamics of quantum-mechanical states of an isolated system is linear and unitary (as in the Schrödinger equation in non-relativistic framework). Second, while in the orthodox interpretation quantum states of systems assign their actual properties, in modal interpretations they assign the *range* of their possible properties and the probabilities of these properties. Third, modal interpretations postulate that systems generally possess more properties than the orthodox interpretation assigns. For example, in contrast to the orthodox interpretation in modal interpretations the systems S, M and O have definite properties in the post-measurement state (3) (for more details, see below).

Different versions of the modal interpretation postulate different property assignments.[6] In the modal interpretations by Kochen (1985), Healey (1989) and

Dieks (1989) (henceforth, the KHD interpretations), the property and probability assignments are based on the *biorthogonal* (Schmidt) decomposition theorem:

KHD Rule Let $\mathcal{H}^{\gamma} = \mathcal{H}^{\alpha} \otimes \mathcal{H}^{\beta}$ be a factorization of the Hilbert space of a composite system γ into the Hilbert spaces of the systems α and β, let $|\Psi\rangle$ be the state of γ, and let the *unique* biorthogonal decomposition of $|\Psi\rangle$ be:

$$|\Psi\rangle = \sum_i \lambda_i |\varphi_i\rangle \otimes |\psi_i\rangle, \tag{4}$$

where $\lambda_i > 0$. Then, α has the property corresponding to the projection $|\varphi_i\rangle\langle\varphi_i|$ and β has the property corresponding to $|\psi_i\rangle\langle\psi_i|$ with the probability $|\lambda_i|^2$.

If the biorthogonal decomposition of $|\Psi\rangle$ is not unique (i.e. if some of the λ_i's are degenerate), the possessed properties are given by the corresponding multi-dimensional projections P_i with probabilities $|\lambda_i|^2 \dim P_i$.[7]

The KHD rule only applies to pure states. But, quantum states of systems can also be represented by *reduced* density operators (obtained by partial tracing). In general, such reduced states are *mixed* states that have no straightforward ignorance interpretation, i.e. they are not proper mixtures. Van Fraassen (1991) proposed to interpret the projections that appear in *any* resolution of a reduced state of a system as representing a subset of its possible properties, one of which is actually possesses. Vermaas and Dieks (1995) make a more restricted choice that relies on the spectral decomposition theorem. Their property and probability assignments are given by the following rule.

Basic Modal Rule Let the reduced state of a system α, associated with the Hilbert space $\mathcal{H}^{\alpha}$, be W and let the spectral resolution of W be:

$$W = \sum_i |\lambda_i|^2 P_i, \tag{5}$$

where P_i are the eigenprojections of W.[8] Then α possesses a property corresponding to P_i with probability $|\lambda_i|^2 \dim(P_i)$.

The Basic Modal Rule selects as definite properties and their probabilities the on-diagonal elements of W. If the spectral resolution of W at some time t is not given in terms of one-dimensional projections, the set of definite properties is still unique, but the definite properties will be non-maximal. The Basic Modal Rule is a generalization of the KHD rule, in that both rules prescribe the same range of properties and their probabilities to any pair of distinct systems (i.e. systems that are associated with non-overlapping Hilbert spaces) in a pure state. Like the KHD rule, the Basic Modal Rule can also be applied to subsystems of composite systems. However, unlike the KHD Rule, the Basic Modal Rule does not assign joint probabilities to the properties of subsystems. In order to assign such probabilities and to account for correlations between properties of different systems, Vermaas and Dieks proposed

the following rule:

Joint Probabilities Let $\mathcal{H}^1 \otimes \ldots \otimes \mathcal{H}^N$ be a factorization of the Hilbert state of a composite system α into the Hilbert spaces of distinct systems $1, 2, \ldots, N$. Let W be α's reduced state, and let the systems $1, \ldots, N$ have reduced states $W^1, \ldots, W^N$, with eigenprojections $\{P^1_{i_1}\}, \ldots, \{P^N_{i_N}\}$, respectively. Then, the joint probability that 1,2, . . . ,N possess the properties $P^1_{i_1}, \ldots, P^N_{i_N}$, respectively, is given by

$$Prob(P^1_{i_1}, \ldots, P^N_{i_N}) = \mathrm{Tr}(WP^1_{i_1} \ldots P^N_{i_N}). \tag{6}$$

Note that Joint Probabilities only assigns probabilities to the properties of distinct systems. Note also that this rule returns the single-time probabilities prescribed by the Basic Modal Rule, the one-to-one correlations implied by the KHD rule and the predictions prescribed by the Born rule for outcomes of joint ideal measurements.

In both the KHD and the Vermaas-Dieks interpretations, the preferred bases are determined by the initial quantum state and the Schrödinger evolution, and therefore they change deterministically over time (i.e. with the evolution of the quantum state). But this is not necessary. Indeed, in Bub's (1992,1997) modal interpretation the preferred bases are time-independent. The definite properties of a system are given by the nonzero projections of its quantum-mechanical state onto the eigenspaces of preferred observables. These observables are distinguished from other observables in that their behavior is stable under decoherence interactions of macroscopic systems with their environment.

3.2 The challenges

Current modal interpretations face three main challenges. First, the property and probability assignments of all current modal interpretations are based on preferred bases. In the KHD and the Vermaas-Dieks interpretations, these are respectively the Schmidt bases and the bases singled out by the spectral resolution. While these bases are selected naturally on mathematical grounds, the question is whether it is possible to justify this selection on physical or metaphysical grounds. Furthermore, as it turns out, these preferred bases are highly unstable in decoherence circumstances, and accordingly the KHD and the Vermaas-Dieks interpretations fail to account for the apparent classical-like behavior of macroscopic systems in such circumstances (see Section 4). And while in Bub's interpretation the choice of the preferred bases is motivated by physical considerations, namely by the stability of the values of observables in experimental circumstances, this choice seems *ad hoc* in that it achieves stability by brute force.

Second, the Basic Modal Rule and the KHD Rule both violate the following conditions concerning the relations between properties of composite systems and the

properties of their subsystems:

> Let $\mathcal{H}^{\alpha}$ and $\mathcal{H}^{\beta}$ be the Hilbert spaces of two distinct systems, α and β. Let P be a projection operator in $\mathcal{H}^{\alpha}$ and let I be the identity operator for $\mathcal{H}^{\beta}$. Then:
>
> *Property Composition.* If α has the property (associated with) P, then $\alpha + \beta$ has the property (associated with) $P \otimes I$.
>
> *Property Decomposition.* If $\alpha+\beta$ has the property (associated with) $P\otimes I$, then α has the property (associated with) P.

In the KHD and Vermaas-Dieks interpretations, the violation of Property Composition and Property Decomposition is necessary for circumventing Kochen&Specker-like no-go theorems (see Bacciagaluppi 1995 and Clifton 1996). Yet, with the exception of Kochen's perspectivalist interpretation, this violation seems inexplicable[9]: While the properties assigned in these interpretations are intrinsic, the violation of Property Composition and Property Decomposition suggests the opposite.

Finally, as the no-go theorems by Dickson and Clifton (1998), Arntzenius (1998) and Myrvold (2002) demonstrate, all current modal interpretations are not genuinely relativistic.

4 THE MEASUREMENT PROBLEM IN MODAL INTERPRETATIONS

The KHD and Vermaas-Dieks interpretations fail to solve the measurement problem. Let us briefly reiterate why. (For the sake of brevity, we shall focus on the Vermaas-Dieks interpretation. But, as is easily shown, a similar analysis holds for the KHD interpretations.)

Consider, again, the final state (3) in the simple measurement scheme presented in Section 2:

$$|\Psi_2\rangle = \lambda_1|\varphi_1\rangle_S|\psi_1\rangle_M|\Phi_1\rangle_O + \lambda_2|\varphi_2\rangle_S|\psi_2\rangle_M|\Phi_2\rangle_O, \tag{7}$$

where, as before, S, M and O denote respectively the measured system, the measuring apparatus and the observer's perception mechanism; $|\varphi_1\rangle_S$ and $|\varphi_2\rangle_S$ are z-spin eigenstates; $|\psi_1\rangle_M$ and $|\psi_2\rangle_M$ are the apparatus pointer eigenstates, corresponding to z-spin 'up' and z-spin 'down' outcomes, respectively; and $|\Phi_1\rangle_O$ and $|\Phi_2\rangle_O$ are eigenstates of a brain observable O, associated with the perception of z-spin 'up' and the perception of z-spin 'down' outcome, respectively. The reduced state of M, W_M, obtained by partial tracing is diagonal in the pointer basis. Accordingly, (assuming that (7) is not a degenerate state) in the Vermaas-Dieks interpretation M points to either 'up' or 'down' with the probabilities $|\lambda_1|^2$ and $|\lambda_2|^2$, respectively. Similarly, the reduced state of O, obtained by partial tracing, is diagonal in the basis of O's eigenstates, and accordingly O has a brain property corresponding to either perceiving 'up' or perceiving 'down'. Furthermore, it follows from Joint Probabilities that

the actual properties of M and O are perfectly correlated. Thus, the Vermaas-Dieks interpretation seems to solve the measurement problem in its basic formulation.

However, the above measurement scheme is highly restricted. More general and realistic models of measurement ought to take into account disturbances and imperfections in the coupling between S and M, and the decoherence interaction of M with its environment E. In such models, the post-measurement state of $S + M + O + E$, expressed in the pointer basis, has generally the form:

$$|\Psi^*(t)\rangle = \sum_j \lambda_j(t)|\varphi_j^*(t)\rangle_S|\psi_j\rangle_M|\Phi_j^*(t)\rangle_O|E_j(t)\rangle_E, \tag{8}$$

where $\{|\varphi_j^*(t)\rangle_S\}$, $\{|\Phi_j^*(t)\rangle_O\}$ and $\{|E_j(t)\rangle_E\}$ are generally sets of non-orthogonal states. According to the standard models of decoherence, the $|E_j(t)\rangle_E$ become in extremely short times approximately orthogonal. Let $W_M(t)$ be the reduced state of M, i.e. the density operator of M obtained by partial tracing from $|\Psi^*(t)\rangle$. In the pointer basis ($\{|\psi_j\rangle_M\}$), $W_M(t)$ has on-diagonal elements of the form $|\lambda_n(t)|^2$ and non-zero off-diagonal elements of the form:

$$\lambda_{ij}(t) := \overline{\lambda}_j(t)\lambda_i(t)\langle\varphi_j^*(t)|\varphi_i^*(t)\rangle_S\langle\Phi_j^*(t)|\Phi_i^*(t)\rangle_O\langle E_j(t)|E_i(t)\rangle_E. \tag{9}$$

Thus, the reduced state of M is not exactly diagonal in the pointer basis, and the properties of M selected by the Basic Modal Rule do not generally correspond to projections onto the pointer eigenstates. Indeed, it has been shown by Bacciagaluppi and Hemmo (1996) that in some cases, e.g. discrete and low-dimensional models of measurement (see Zurek 1991), the properties of M correspond to projections onto states that are pairwise *close* in Hilbert space norm to the pointer eigenstates if $|\Psi^*(t)\rangle$ is far enough from degeneracy. Bacciagaluppi and Hemmo argue that these properties may naturally be interpreted as genuinely close to the properties corresponding to definite pointer readings. But, in other cases, e.g. in continuous models of measurement with position being the pointer observable, the reduced state of M, $W_M(t)$, becomes *extremely* degenerate as a result of its decoherence interaction with the environment. Consequently, the on-diagonal elements of $W_M(t)$ are projections that correspond to highly *delocalized* wavefunctions (see Joos and Zeh 1985, Bacciagaluppi and Hemmo 1996 and Bacciagaluppi 2000). Thus, in these general models of measurement, pointers have no definite positions, and even worse the Bacciagaluppi-Hemmo strategy fails; for the distance between the projections that appear in the spectral resolution of $W_M(t)$ and the projections onto definite positions turns out to be very substantial. And so the KHD and the Vermaas-Dieks modal interpretations fail to solve the measurement problem in these more general and realistic models of measurement.

In addition to the measurement problem, the KHD and the Vermaas-Dieks interpretations also face another challenge in accounting for our perception of, and beliefs about the classical-like behavior of macroscopic systems. Recall (Section 3.2) that these interpretations violate Property Composition and Property Decomposition. This means that while properties are assigned to systems in every partition of the universe

into subsystems, properties that are assigned in different partitions are generally *unrelated* to each other. Thus, similarly to any other system, the observables of observers' brains may be assigned different values in different partitions. The question is then: How are all these different properties related to our beliefs about the physical systems that appear in our experience? It may be tempting to postulate Property Composition and Property Decomposition. But, recalling (Section 3.2) the Kochen&Specker-like theorems by Bacciagaluppi (1995) and Clifton (1996), such postulation will lead to inconsistency.

5 THE RELATIONAL MODAL INTERPRETATION

In the next three sections we introduce the basic ideas and postulates of the relational modal interpretation. In this section, we introduce a non-relativistic version of it. This version provides an explanation for the failure of Property Composition and Property Decomposition, abolishes the preferred-basis property assignment of current modal interpretations, and addresses the problems that the KHD and Vermaas-Dieks interpretations encounter in non-ideal measurements and realistic models of decoherence. In Section 6, we discuss the way the relational modal interpretation accounts for our perception of the classical-like behavior of macroscopic systems. In Section 7, we revise the dynamics of properties and in Section 8 we demonstrate how due to the radical nonlocal nature of this revised dynamics the relational interpretation circumvents the no-go theorems for relativistic modal interpretations. In Section 9, we explain how the relational modal interpretation accounts for our failure to perceive this radical type of nonlocality. Finally, in Section 10 we consider the prospects of reconciling this interpretation with special relativity.

5.1 The property assignment

In the relational modal interpretation, the general idea of the property assignment is that reduced states of systems do not assign the range of their possible intrinsic properties but rather the range of their possible relational properties. That is, let α and β be any partition of the universe into two distinct systems. Then, at any time t the on-diagonal elements of the reduced state of α in every orthonormal resolution provides a set of (mutually exclusive) properties that α may have *relative* to β and the single-time probabilities of these properties at t. The basic postulates of this property and probability assignments are as follows.

Relational Property Rule Let $\mathcal{H} = \mathcal{H}^{\alpha} \otimes \mathcal{H}^{\beta}$ be a factorization of the Hilbert space of the universe into the Hilbert spaces of two distinct systems, α and β. Let the reduced state of α (obtained by tracing the state of $\alpha + \beta$ over $\mathcal{H}^{\beta}$) be W_{α}, and let $\{Q_i\}$ be *any* orthonormal basis of $\mathcal{H}^{\alpha}$. Then, the set of non-zero projections

$$\{W_{\alpha} Q_i\} \tag{10}$$

corresponds to a set of (mutually exclusive) properties that α may have relative to β in the state W, and the single-time probabilities of these properties are given by $\mathrm{Tr}(W_\alpha Q_i)$.

Note that here the properties that α has relative to β are related to β *simpliciter* rather than to β's particular properties. This type of relational properties is in a sense 'thinner' than the relational properties postulated by Everett-type theories, where the properties of α are related to particular *properties* of β.

Properties that are related to the same systems (context) have joint probabilities in accordance with the following decomposition rule.

Relational Decomposition Rule Let $\mathcal{H} = \mathcal{H}^\alpha \otimes \mathcal{H}^\beta$ be a factorization of the Hilbert space of the universe into the Hilbert spaces of two distinct systems, α and β, and let $\mathcal{H}^\alpha = \mathcal{H}^{\alpha^1} \otimes \mathcal{H}^{\alpha^2}$ be a factorization of the Hilbert space of α into the Hilbert spaces of two distinct systems, α_1 and α_2. Let the reduced state of α be W_α, and let $\{P_i\}$ and $\{Q_i\}$ be respectively orthonormal bases in the Hilbert spaces $\mathcal{H}^{\alpha^1}$ and $\mathcal{H}^{\alpha^2}$, such that the on-diagonal elements of W_α correspond to projections that have the product form:

$$P_i \otimes Q_i. \tag{11}$$

Then, as subsystems of α, the single-time probability that α^1 has a property P_i relative to β is $\mathrm{Tr}(W_\alpha P_i)$, the single-time probability that α^2 has a property Q_i relative to β is $\mathrm{Tr}(W_\alpha Q_i)$, and the single-time joint probability of these properties is $\mathrm{Tr}(W_\alpha P_i Q_i)$.

The properties that α^1 and α^2 have relative to β are different from, and cannot be identified with the properties that α^1 has relative to $\alpha^2{+}\beta$ and α^2 has relative to $\alpha^1{+}\beta$. Thus, it follows from the Relational Property Rule and the Relational Decomposition Rule that there is no way to assign joint single-time probabilities to properties that are related to different contexts (systems) on the basis of quantum-mechanical states. For according to these rules, there are no reduced states to provide joint single-time probabilities for properties that are related to different contexts. The reasoning is as follows. Consider, for instance, the probability that α^1 has the property P relative to $\alpha^2 + \beta$, the probability that α^2 has the property Q relative to $\alpha^1 + \beta$ and the joint probability of these properties. The probability that α^1 has the property P relative to $\alpha^2 + \beta$ is given by the reduced state of α^1, and the probability that α^2 has the property Q relative to $\alpha^1 + \beta$ is given by the reduced state of α^2. But, the joint probability of these properties is not given by the reduced state of α ($\alpha^1 + \alpha^2$) or any other reduced state. According to the Relational Decomposition Rule, the reduced state of α only gives the joint probability of the properties that α^1 and α^2 each has relative to β, and these probabilities need not be the same as the joint probabilities that α^1 has relative to $\alpha^2{+}\beta$ and α^2 has relative to $\alpha^1{+}\beta$. Thus, assuming that single-time probabilities can only be assigned on the basis of quantum-mechanical states, properties that are related to different contexts have no joint probabilities and accordingly are unrelated to each other.

Similarly, based on this assumption and the assumption that properties that are assigned in different resolutions of a reduced state (i.e. in different bases)

do not constrain each other (an assumption that is required for circumventing Kochen&Specker-like no-go theorems), properties that are assigned in different bases have no joint probabilities and accordingly are uncorrelated with each other. For granted the above assignment rules, there is no way to assign joint probabilities for such properties on the basis of quantum-mechanical states.

As is not difficult to see, the above property assignment does not single out any preferred bases. Further, the failure of Property Composition and Property Decomposition is naturally explained. For this property assignment only provides the range of systems' possible relational properties, and due to their nature properties that a system has relative to different contexts need not be the same.

5.2 The non-relativistic dynamics

We now turn to introduce an outline of the dynamics of relational properties in non-relativistic framework. Let U be a unitary transformation on the state of the composite system $\alpha+\beta$. The range of possible properties that α and its subsystems have relative to β and the single-time probabilities of these properties at any time t, are determined by α's reduced state at t, $W_\alpha(t)$. The dynamics of these properties and their probabilities satisfies the following conditions:

D1 If $W_\alpha(t)$ does not change under U, the range of the possible properties of α (and its subsystems) relative to β and the single-time probabilities of these properties do not change. Further, the actual properties of α (and its subsystems) relative to β also remain unchanged.

D2 If $W_\alpha(t)$ changes under U, the evolution of the properties that α and its subsystems have relative to β depends on U in the following way. (i) Properties of α (and its subsystems) relative to β associated with projections that *commute* with U evolve deterministically, so as to return the single-time Born probabilities. (ii) Properties of α (and its subsystems) relative to β associated with projections that *don't commute* with U evolve indeterministically, so as to return the single-time Born probabilities.

D3 Let α^1 be a subsystem of α, and let $\mathrm{Prob}(Q(t_2)|P(t_1))$ be the transition probability that α^1 has, as a subsystem of α, the property Q relative to β at t_2 *given* that it has, as a subsystem of α, the property P relative to β at time t_1. In general, $\mathrm{Prob}(Q(t_2)|P(t_1))$ depends not only on the properties that α_1 has relative to β but rather on the properties that α has relative to β.

Four remarks about the above outlines of dynamics: First, as is easily seen, D3 implies that the dynamics of properties is holistic in nature. Second, note that D1-D3 do not pick out a unique dynamics; rather, they are compatible with a class of possible dynamics. Third, each of the dynamics in this class reproduces the single-time probabilities obtained by a sequential application of the Born rule in collapse theories, and accordingly recovers all the predictions of standard collapse quantum mechanics. Fourth, as we shall see, these dynamics are special cases of a more universal dynamics we develop in Section 7 – namely, they are dynamics that would surface

when the relevant degree of entanglement is zero, as the case is in states of perfect decoherence.

6 EXPERIENCE IN THE RELATIONAL MODAL INTERPRETATION

We now turn to consider how the relational modal interpretation addresses the measurement problem and, more generally, accounts for our classical-like experience. Let S_I consist of a measured system S, a measuring apparatus M, and two observers O_1 and O_2, and let S_{II} consists of the environment of these systems and the rest of the universe. Suppose that M carries out an ideal z-spin measurement on S, and O_1 and O_2 both observe the measurement outcome. The post-measurement quantum-mechanical state of S_I ($S + M + O_1 + O_2$) and S_{II} has the general form:

$$|\Psi\rangle = \sum_{i,j,k,l,m} \lambda_{i,j,k,l,m} |\varphi_i\rangle_S |\psi_j\rangle_M |\Phi^1_k\rangle_{O_1} |\Phi^2_l\rangle_{O_2} |\xi_m\rangle_{S_{II}}, \tag{12}$$

where $|\varphi_i\rangle_S$ ranges over $|z+\rangle$ (z-spin 'up') and $|z-\rangle$ (z-spin 'down'), $|\psi_j\rangle_M$ ranges over $|up\rangle$ (pointer pointing to 'up') and $|down\rangle$ (pointer pointing to 'down'), and $|\Phi^1_k\rangle_{O_1}$ and $|\Phi^2_l\rangle_{O_2}$ each ranges over 'b−up' (the brain state associated with the state of mind of believing 'up') and 'b−down'(the brain state associated with the state of mind of believing 'down'). There always exists a normalized basis $\{|r_1\rangle, |r_2\rangle\}$ in the Hilbert space associated with S_{II}, such that the state (12) can be rewritten as follows:

$$\begin{aligned} |\Psi\rangle = {} & \lambda_1 (|z+\rangle_S |\mathrm{up}\rangle_\mathrm{M} |\mathrm{b-up}\rangle_{\mathrm{O_1}} |\mathrm{b-up}\rangle_{\mathrm{O_2}}) |\mathrm{r_1}\rangle_{\mathrm{S_{II}}} + \\ & + \lambda_2 (|z-\rangle_S |\mathrm{down}\rangle_\mathrm{M} |\mathrm{b-down}\rangle_{\mathrm{O_1}} |\mathrm{b-down}\rangle_{\mathrm{O_2}})_{\mathrm{S_1}} |\mathrm{r_2}\rangle_{\mathrm{S_{II}}} \end{aligned} \tag{13}$$

where $|r_1\rangle_{S_{II}}$ and $|r_2\rangle_{S_{II}}$ are not necessarily orthogonal. Since the properties of S_I (i.e. of $S + M + O_1 + O_2$) relative to S_{II} are given by S_I's reduced state, it follows that S, M, O_1 and O_2 have definite relational properties that are appropriately correlated with each other. Suppose, for example, that relative to S_{II} the system S has (as a subsystem of S_I) the property z-spin 'up'. Then, relative to S_{II} the position of the apparatus's pointer (as a subsystem of S_I) is 'up' and the brain properties of both observers (as subsystems of S_I) are those of 'b−up' (i.e. the brain state associated with believing pointer 'up').

This analysis of experience can be generalized to any number of systems and observers. Further, this schematic analysis of ideal measurements can easily be generalized to account for models of non-ideal measurements. For even in non-ideal measurements, where pointer observables do not get perfectly correlated with the measured observables, there are always bases in which the properties of S, M, O_1 and O_2 relative to S_{II} (assigned by S_I's reduced state in the post-measurement state (13)) correspond to our classical-like experience, regardless of any decoherence interaction of S_I (or some of its subsystems) with the environment in S_{II}. For example, the projections onto the pointer states, $|\mathrm{up}\rangle_\mathrm{M}$ and $|\mathrm{down}\rangle_\mathrm{M}$ in state (13) correspond

to definite properties of M relative to S_{II} with the respective probabilities $|\lambda_1|^2$ and $|\lambda_2|^2$, regardless of the size of the off-diagonal elements of the reduced state of M in the pointer basis. Thus, there always exist subsets of relational properties that may correspond to our perception of definite pointer readings; and this is true even when decoherence is not sufficiently effective to diagonalize the reduced state of S_I in (13). Further, one can show that in approximate decoherence the behavior of these properties is stable.

The above analysis may also account for our belief in the existence of systematic correlations between the properties of physical systems and our beliefs about them. It is commonly assumed that there exist systematic correlations between certain brain properties and states of mind. If we assume by analogy that such correlations exist between observers' brain properties that are related to a certain context and their states of mind, then it is possible to show that there exist systematic correlations between subsets of relational properties of physical systems and observers' beliefs about these systems. Yet, since subsets of properties that are related to different contexts (and likewise subsets of properties that are assigned in different bases) are unrelated, the question arises as to which subsets of relational properties are correlated with our beliefs about, and our experience of the physical world.

We believe that the question of the exact nature of the brain properties that are related to observers' experience and beliefs is not unique to the relational modal interpretation. It also arises in other interpretations. But, since the relational interpretation postulates the existence of many uncorrelated subsets of definite properties, this question seems to be more acute in the context of this interpretation. In the context of the current interpretation, it is plausible to assume that experience is associated with a subset of relational properties that are related to a single context.[10] Further, it seems also plausible to assume that the identity of this subset of physical properties (partially) depends on decoherence interactions with the environment. Yet, we believe that in the state of current knowledge about the relationships between our experience and the teachings of contemporary physics, the best one could do is to give schematic models in the context of which it is possible to tell an intelligible story about how our experience of the physical world may be reconciled with a given interpretation of quantum mechanics. In particular, one may be able to demonstrate in such models that there exist systematic correlations between some subsets of physical properties of systems, physical properties of observers' brains and our mental states and beliefs about the physical world. As we have argued above, the relational modal interpretation may well have such an intelligible story to tell.

7 THE UNIVERSAL DYNAMICS

The dynamics outlined in D1-D3 of Section 5 is subjected to several no-go theorems for relativistic modal interpretations (for a discussion of these theorems, see Section 8). In this section, we propose that this dynamics is a special case of a more universal dynamics that circumvents these theorems. This universal dynamics may be introduced as a sum average over two extreme cases: (i) the dynamics in case of

no entanglement; and (ii) the dynamics in case of maximal entanglement. We shall define below the relevant notion of entanglement and its measure. But first we turn to introduce the dynamics in these extreme cases.

The dynamics in case of no entanglement is similar to the dynamics outlined in Section 5. That is, it is the dynamics outlined in D1-D3 with the required modifications for a relativistic framework. Instead of being formulated with reference to states at different times, transition probabilities are now formulated with reference to states on different spacelike hypersurfaces: The probability that a system has a relational property Q on a spacelike hypersurface σ_2 (i.e. in the state that obtains on σ_2) given that it has a relational property P on a spacelike hypersurface σ_1 (i.e. in the state that obtains on σ_1).

The dynamics in the case of maximal entanglement can be expressed by the following condition.

D4 Let $\alpha + \beta$ be a partition of the universe into two subsystems; let $W_\alpha(\sigma_1)$ and $W_\alpha(\sigma_2)$ be the reduced states of α on the spacelike hypersurfaces σ_1 and σ_2, respectively; and let Q and P be any two properties that α may have relative to β. Then, if $W_\alpha(\sigma_2) \neq W_\alpha(\sigma_1)$, the transitional probability $\text{Prob}_{\text{me}}(Q|P)$, namely the probability that α has the property Q relative to β in the state $W_\alpha(\sigma_2)$ given that it has the property P relative to β in the state $W_\alpha(\sigma_1)$, is equal to the single-time Born probability that α has the property Q relative to β in the state $W_\alpha(\sigma_2)$:

$$\text{Prob}_{\text{me}}(Q|P) = Tr(W_\alpha(\sigma_2)Q). \tag{14}$$

If $W_\alpha(\sigma_2) = W_\alpha(\sigma_1)$, then the properties that α has relative to β do not change.

The notion of entanglement we shall work with is bipartite, i.e. it applies to the entanglement between two systems. The *measure* of entanglement between two systems, α and β, in a state $|\psi\rangle$ may be defined as the minimal (normalized) distance, in Hilbert space norm, between $|\psi\rangle$ and all the possible product states in the Hilbert space $\mathcal{H}^\alpha \otimes \mathcal{H}^\beta$ (see, for example, Shimony 1995). But other measures of entanglement may also be applicable. This geometrical measure of entanglement may easily be generalized to *mixed* states. In that case, the measure of entanglement is defined as the minimal (normalized) distance between the mixed state of systems and the set of all their product states, as follows. Let W be any mixed state in $\mathcal{H}^\alpha \otimes \mathcal{H}^\beta$, and let $\mathbf{C}$ be the convex set of all the mixed product states of α and β. The degree of entanglement between α and β is defined as the normalized distance between W and any state $W_i \in \mathbf{C}$, such that $\text{Tr}(W \otimes W_i) \leq \text{Tr}(W \otimes W_j)$ for all states $W_j \in \mathbf{C}$.[11]

In the context of the relational modal interpretation, the relevant measure of entanglement depends on the relational properties under consideration and the transformations of the quantum state. Let α and β be a partition of the universe, α^1 and α^2 be a partition of α, and β^1 be a subsystem of β. Let $U(\sigma_1, \sigma_2)$ be any (non-identity) unitary transformation of the quantum state of $\alpha^1 + \beta^1$ and the identity transformation of the state of all the other systems of the universe from a spacelike hypersurface σ_1

to a spacelike hypersurface σ_2. Then, if neither α^1 nor α^2 is the 'null' system, the transition probabilities of the properties of α and its subsystems relative to β under the transformation $U(\sigma_1, \sigma_2)$ depend on the degree of entanglement between α^2 and $\alpha^1 + \beta^1$ in the (reduced) state of $\alpha^1 + \alpha^2 + \beta^1$ (i.e. $\alpha + \beta^1$) on σ_1. If α^1 or α^2 is the 'null' system, the degree of entanglement is postulated to be zero. This measure of entanglement is supposed to reflect the extent to which the properties of α relative to β have to be redistributed, so as to reproduce the Born probabilities on every spacelike hypersurface without picking out any preferred foliation of spacetime.

The universal dynamics is a weighted average of the dynamics in the two extreme cases of no entanglement and of maximal entanglement. That is, let (as before) $U(\sigma_1, \sigma_2)$ be any unitary (non-identity) transformation on the state of $\alpha^1 + \beta^1$ and the identity transformation on the state of all other subsystems of $\alpha + \beta$ from a spacelike hypersurface σ_1 to a spacelike hypersurface σ_2, and let $|\psi(\sigma_1)\rangle$ be the state of $\alpha + \beta$ on σ_1. Then, the conditional probability that α has the property Q relative to β in the state $|\psi(\sigma_2)\rangle$ $(= U(\sigma_1, \sigma_2)|\psi(\sigma_1)\rangle)$ given that it has the property P relative to β in the state $|\psi(\sigma_1)\rangle$, $\mathrm{Prob}_U(Q|P)$, is given by:

$$\mathrm{Prob}_U(Q|P) = d(e) \cdot \mathrm{Prob}_{\mathrm{me}}(Q|P) + \tag{15}$$

$$+ (1 - d(e)) \cdot \mathrm{Prob}_{\mathrm{ne}}(Q|P);$$

where $\mathrm{Prob}_{\mathrm{me}}(Q|P)$ and $\mathrm{Prob}_{\mathrm{ne}}(Q|P)$ are the conditional probabilities of Q given P according to the dynamics in case of maximal entanglement and the dynamics in case of no entanglement respectively, and $d(e)$ is the degree of entanglement between $\alpha^1 + \beta^1$ and α^2 in the state $|\psi(\sigma_1)\rangle$. If the distribution of properties of α relative to β is given by the single-time Born-like probabilities on any spacelike hypersurface, then (by construction) the transition probabilities in both maximal entanglement and no entanglement reproduce the single-time Born probabilities on any other spacelike hypersurface. Since the universal dynamics is a weighted average of the dynamics in these extreme cases, it similarly reproduces the Born probabilities.[12]

8 THE NO-GO THEOREMS FOR RELATIVISTIC MODAL INTERPRETATIONS

Bell's theorem suggests that given natural assumptions, any adequate quantum theory will have to postulate some non-local influences.[13] In contrast to a popular view, not all non-local influences imply incompatibility with special relativity.[14] Special relativity requires that the dynamical laws of properties do not select any preferred inertial reference frame (i.e. preferred foliation of spacetime into parallel spacelike hyperplanes) and that the description of systems' properties and their probabilities in different frames are compatible with each other. Thus, the question of the compatibility of an interpretation of quantum mechanics with relativity turns on whether the dynamics of properties satisfies these constraints and reproduces Born-like probabilities for the possessed properties of systems for arbitrary initial state along every inertial

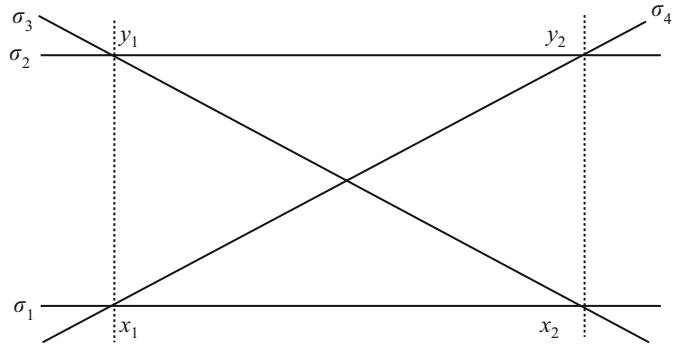

FIGURE 1.1. The spacelike hypersurfaces involved in Myrvold's theorem.

reference frame. The no go theorems by Dickson and Clifton (1998), Arntzenius (1998) and Myrvold (2002) demonstrate that current modal interpretations fail to do so, and accordingly a common view has it that these interpretations cannot be made genuinely relativistic. In the next two subsections, we shall consider Myrvold's and Dickson and Clifton's theorems and show why they do not apply to the relational modal interpretation. For want of space, we shall not be able to discuss Arntzenius's theorem. But based on the discussion of Myrvold's theorem, it is not difficult to show that Arntzenius's theorem also fails to apply to this interpretation.

8.1 Myrvold's theorem

Myrvold's (2002) theorem asserts that in the property assignments of current modal interpretations (see Section 3.1), the probabilities of local possessed properties cannot be given by the Born probabilities along every foliation of spacetime for arbitrary initial quantum state, irrespectively of their dynamics.[15]

Myrvold considers the following set up (see Figure 1.1). σ_1 and σ_2 are two hyperplanes of simultaneity in some reference frame. x_i (y_i) is a small region on σ_1 (σ_2) in which a system S_i is located. x_1 is spacelike separated from y_2 and x_2 is spacelike separated from y_1. σ_3 is a spacelike hypersurface containing y_1 and x_2, and σ_4 is a spacelike hypersurface containing x_1 and y_2. R_1 and R_2 are observables of the systems S_1 and S_2 respectively, coupled to measuring devices A_1 and A_2 respectively, which record the values of R_1 and R_2 on $\sigma_1 - \sigma_4$. The states of $S_1 + S_2 + A_1 + A_2$ on these hypersurfaces are:

$$|\varphi(\sigma_1)\rangle = 1/2\sqrt{3}(|p_1+\rangle|r_1+\rangle|r_2+\rangle|p_2+\rangle - |p_1+\rangle|r_1+\rangle|r_2-\rangle|p_2-\rangle - |p_1-\rangle|r_1-\rangle|r_2+\rangle|p_2+\rangle - 3|p_1-\rangle|r_1-\rangle|r_2-\rangle|p_2-\rangle); \quad (16)$$

$$|\varphi(\sigma_2)\rangle = 1/\sqrt{3}(|p_1+\rangle|r_1+\rangle|r_2-\rangle|p_2-\rangle + |p_1-\rangle|r_1-\rangle|r_2+\rangle|p_2+\rangle - |p_1+\rangle|r_1+\rangle|r_2+\rangle|p_2+\rangle); \quad (17)$$

$$|\varphi(\sigma_3)\rangle = 1/\sqrt{6}(|p_1-\rangle|r_1-\rangle|r_2+\rangle|p_2+\rangle + |p_1-\rangle|r_1-\rangle|r_2-\rangle|p_2-\rangle - \\ - 2|p_1+\rangle|r_1+\rangle|r_2-\rangle|p_2-\rangle); \quad (18)$$

$$|\varphi(\sigma_4)\rangle = 1/\sqrt{6}(|p_1+\rangle|r_1+\rangle|r_2-\rangle|p_2-\rangle + |p_1-\rangle|r_1-\rangle|r_2-\rangle|p_2-\rangle - \\ - 2|p_1-\rangle|r_1-\rangle|r_2+\rangle|p_2+\rangle); \quad (19)$$

where for each i, $|r_i+\rangle$ and $|r_i-\rangle$ are distinct eigenstates of the observable R_i; and $|p_i+\rangle$ and $|p_i-\rangle$ are distinct eigenstates of P_i, a pointer observable of the measuring device A_i. As is easily seen, $|\varphi(\sigma_2)\rangle$, $|\varphi(\sigma_3)\rangle$ and $|\varphi(\sigma_4)\rangle$ are obtained from $|\varphi(\sigma_1)\rangle$ by applying the following Hadamard transformation to the eigenstates of $R_i \otimes P_i$:

$$U_i|r_i+\rangle|p_i+\rangle = 1/\sqrt{2}(|p_i+\rangle|r_i+\rangle + |p_i-\rangle|r_i-\rangle); \quad (20)$$

$$U_i|r_i-\rangle|p_i-\rangle = 1/\sqrt{2}(|p_i+\rangle|r_i+\rangle - |p_i-\rangle|r_i-\rangle).$$

That is, $|\varphi(\sigma_2)\rangle = U_1 \otimes U_2|\varphi(\sigma_1)\rangle$, $|\varphi(\sigma_3)\rangle = U_1 \otimes I_2|\varphi(\alpha)\rangle$ and $|\varphi(\sigma_4)\rangle = I_1 \otimes U_2|\varphi(\alpha)\rangle$; where I is the identity transformation.

The main idea of Myrvold's theorem is as follows. In the Schmidt-decomposition and the spectral-resolution modal interpretations, R_1 and R_2 have definite values on $\sigma_1 - \sigma_4$. Suppose that the same is true for Bub's modal interpretation. Further, suppose that the values of R_i correspond to local properties. Then, these values are the same on any two space-like hypersurfaces that intersect the spacetime region in which S_i is located. Moreover, if the probabilities of these values were to satisfy the Born rule for single-time probabilities on the hypersurfaces $\sigma_1 - \sigma_4$, there would have to be a joint probability distribution over these values that yields as marginals their Born probabilities on all the four hypersurfaces. But, such joint probability distribution would satisfy certain Bell-type inequalities, which are violated in states $|\varphi(\sigma_1)\rangle - |\varphi(\sigma_4)\rangle$ and various other states (for more details, see Myrvold 2002 and Berkovitz and Hemmo 2005a). This means that the probabilities of such local properties cannot be given by the Born probabilities along every foliation of spacetime into parallel spacelike hyperplanes for an arbitrary initial quantum state.

Consider, for example, the values that R_1 and R_2 have on the hypersurfaces σ_1 - σ_4 in the set up of Myrvold's theorem, i.e. in the states (16)-(19).[16] Suppose that on σ_1 R_1 and R_2 have the values r_1+ and r_2+, respectively. By assumption, R_1 is a local property of S_1, and R_2 is a local property of S_2. Thus, R_1 must have the same value on the hypersurface σ_4, and accordingly it follows from the Born rule that the probability that R_1 and R_2 have, respectively, the values r_1+ and r_2- on σ_4 is one. Further, the value of R_2 on σ_2 is the same as its value on σ_4. Accordingly, the probability that R_2 has the value r_2- on σ_2 given that R_1 and R_2 have the values r_1+ and r_2+ on σ_1 is one. A parallel argument leads to the conclusion that if R_1 and R_2 have the values r_1+ and r_2+ on σ_1, the probability that R_1 has the value r_1- on σ_2 is also one. Thus, if R_1 and R_2 have the values r_1+ and r_2+ on σ_1, the probability they have the values r_1- and r_2- on σ_2 is one. But, by the Born rule $|\varphi(\sigma_2)\rangle$ assigns zero probability to these values.

Myrvold's theorem rests on the following three premises:

(i) *Local Properties.* There exist local observables, R_i ($i = 1, 2$), the value of which is the same on any two spacelike hypersurfaces that intersect the region in which the system they pertain to, S_i, is located.

(ii) *Joint Probabilities.* There exist joint probabilities for the values of R_1 and R_2 on the hypersurfaces σ_1 - σ_4, which yield as marginals the Born probabilities for the values of R_1 and R_2 on all these four hypersurfaces.

(iii) *Relativistic Born Rule.* The joint probabilities of local properties on every spacelike hypersurface are given by the Born probabilities. That is, let q and r be any possible values of the observables Q_1 and R_2 respectively, and let $Q_1 = q$ and $R_2 = r$ be local definite properties of the systems S_1 and S_2, respectively. For any spacelike hypersurface σ, if the quantum-mechanical state of the composite system $S_1 + S_2$ on σ is $|\psi(\sigma)\rangle$, then the probability of $Q_1 = q$ and $R_2 = r$ on σ is equal to $\mathrm{Tr}[P_{Q_1}(q)P_{R_2}(r)|\psi(\sigma)\rangle]$; where $P_{Q_1}(q)$ and $P_{R_2}(r)$ are the projections onto the eigenspaces $Q_1 = q$ and $R_2 = r$, respectively.[17]

But the relational modal interpretation violates these conditions. Here is why. Consider the value that R_1 has, as a property of S_1, relative to $S_2 + A_1 + A_2$ and the value that R_2 has, as a property of S_2, relative to $S_1 + A_1 + A_2$. These values are related to different contexts. Thus, it follows from the property assignment of this interpretation that they have no definite joint probabilities, and accordingly Myrvold's theorem does not apply to them.

Consider now the values that R_1 and R_2 have, as properties of $S_1 + S_2$, relative to $A_1 + A_2$. In the relational modal interpretation, these values are definite on $\sigma_1 - \sigma_4$ (i.e. in the states that obtain on these spacelike hypersurfaces). Further, it follows from the Relational Decomposition Rule that they have joint probabilities on all the four hypersurfaces. But, by the universal dynamics, Local Properties cannot be assumed: The probability that the value of R_1 (R_2), as a property of $S_1 + S_2$, relative to $A_1 + A_2$ on σ_1 will be different from its value on σ_4 (σ_3) is proportional to the degree of entanglement between S_1 and $S_2 + A_2$ (S_2 and $S_1 + A_1$) on σ_1. The higher the degree of entanglement is, the higher is the probability that this value of R_1 (R_2) will not be the same on σ_1 and σ_4 (σ_3). Similarly, the probability that the value that R_1 (R_2) has, as a property of $S_1 + S_2$, relative to $A_1 + A_2$ will not be the same on σ_3 and σ_2 (σ_4 and σ_2) is proportional to the degree of entanglement between S_1 and $S_2 + A_2$ (S_2 and $S_1 + A_1$) on σ_3 (σ_4). Since these degrees of entanglement are substantial, Local Properties cannot be assumed. Accordingly Myrvold's theorem is inapplicable to the values that R_1 and R_2 have, as properties of $S_1 + S_2$, relative to $A_1 + A_2$.

Note that the dependence of the universal dynamics on the degree of entanglement is desirable. It circumvents Myrvold's theorem by yielding radical non-locality in non-experimental circumstances, where the (relevant) degree of entanglement is significant. Yet, as we shall see in Section 9, this non-locality is unobservable in experimental circumstances, where the (relevant) degree of entanglement is virtually zero. Note also that it is not the relational nature of the values of R_1 and R_2 *per se* that are 'responsible' for circumventing Myrvold's theorem. Indeed, by their very nature the properties postulated by the relational modal interpretation are nonlocal: They are

relations between distant systems. But this nonlocality is not sufficient for circumventing Myrvold's theorem, as some relational properties behave like local properties. For example, the value that R_1 has, as a property of S_1, relative to $S_2 + A_1 + A_2$ is highly nonlocal by its very nature. Yet, as the relevant degree of entanglement for the dynamics of this value is zero, the universal dynamics dictates that it is the same on σ_1 and σ_4; and similarly, *mutatis mutandis*, for the value of R_2, as a property of S_2, relative to $S_1 + A_1 + A_2$ on σ_1 and σ_3. Accordingly, these values behave like local properties, and Myrvold's theorem is inapplicable to them because of the failure of Joint Probabilities rather than their nonlocal nature. By contrast, the relevant degrees of entanglement for the dynamics of the values that R_1 and R_2 have, as properties of $S_1 + S_2$, relative to $A_1 + A_2$ on σ_1 are significant, and accordingly the probability that these relational values are not the same on σ_1 and σ_4 (σ_3) is substantial. This further type of nonlocality (henceforth, *e-nonlocality*) is what renders Myrvold's theorem inapplicable to these relational values of R_1 and R_2.

Generalizing the above reasoning, it is not difficult to show that Myrvold's theorem is also inapplicable to other relational values of R_1 and R_2 and, more generally, any other relational properties postulated by the relational modal interpretation.

8.2 Dickson and Clifton's theorem

In their theorem, Dickson and Clifton (1998) demonstrate that granted certain premises motivated by relativistic and dynamical considerations, the KHD and the Vermaas-Dieks modal interpretations fail to be genuinely relativistic. In reference to the Einstein-Podolsky-Rosen/Bohm (EPR/B) experiment, they consider the probabilities that the particles possess certain spin properties before and after the spin measurements in three different foliations of spacetime, which can be associated with three different inertial reference frames: The reference frame S in which the measurements occur simultaneously; the reference frame L in which the left-hand-side measurement occurs first; and the reference frame R in which the right-hand-side measurement occurs first.

Dickson and Clifton's theorem relies on four main premises. The first two premises are motivated by relativistic considerations.

(i) *Fundamental Lorentz Invariance.* When a system undergoes a free evolution, its definite properties in different inertial frames are related by the Lorentz transformations.

(ii) *Invariant Transition Probabilities.* When a system undergoes a free evolution, the transition probabilities of its properties in different inertial frames are related by the Lorentz transformations.

The third premise is motivated by dynamical considerations.

(iii) *Stability.* In any frame of reference, if no measurement is made on a system in the time interval $[t_1, t_2]$ $(t_1 < t_2)$, the probability that the system has the property P at time t_2 given that it has that property at t_1 is one.[18]

Finally, Dickson and Clifton also presuppose:

(iv) *Joint Probabilities.* The spin properties of the particles in the EPR/B experiment have definite joint probabilities.

Dickson and Clifton demonstrate that modal interpretations that satisfy the above four conditions cannot reproduce the predictions of orthodox quantum mechanics in the reference frames S, L and R. In the relational modal interpretation, the corresponding properties are the properties that the L-particle has relative to the R-particle, the measurement apparatuses and the rest of the universe and the spin properties that the R-particle has relative to the L-particle, the measurement apparatuses and the rest of the universe. For these relational properties, the relevant degrees of entanglement are zero. Thus, their dynamics is in effect the dynamics in case of no entanglement, where Fundamental Lorentz Invariance, Invariant Transition Probabilities and Stability hold. But, since the properties of the L-particle and the R-particle are related to different contexts, they do not have joint probabilities. Accordingly, Joint probabilities fails and Dickson and Clifton's theorem does not apply to these properties.

9 WHY E-NONLOCALITY IS UNOBSERVABLE

In the relational modal interpretation, the nature of nonlocality depends on entanglement. When the relevant degree of entanglement is nonzero, the dynamics of apparently local quantities, such as the value of R_1 as a property of $S_1 + S_2$ relative to $A_1 + A_2$, may involve e-nonlocality (see Section 8.1): This value of R_1 may not be the same on two spacelike hypersurfaces that intersect the region in which S_1 is located. Yet, our experience seems to suggest that this type of non-locality never occurs. So the question is why it is unobservable.

To answer this question, let us consider again Myrvold's set up (see Section 8.1), but now suppose that the measuring apparatuses A_1 and A_2 are subjected to decoherence interactions with their local environment E. In the ideal case, the state of $S_1 + S_2 + A_1 + A_2 + E$ will be

$$\begin{aligned} |\varphi(\sigma_1)\rangle = 1/2\sqrt{3}(&|p_1+\rangle|r_1+\rangle|r_2+\rangle|p_2+\rangle|E++\rangle - \\ &- |p_1+\rangle|r_1+\rangle|r_2-\rangle|p_2-\rangle|E+-\rangle - \\ &- |p_1-\rangle|r_1-\rangle|r_2+\rangle|p_2+\rangle|E-+\rangle - \\ &- 3|p_1-\rangle|r_1-\rangle|r_2-\rangle|p_2-\rangle|E--\rangle, \end{aligned} \tag{21}$$

where the $|E++\rangle, |E+-\rangle, |E-+\rangle$ and $|E--\rangle$ are the orthogonal states of the environment.

Due to the perfect decoherence with the environment, the degree of entanglement between S_1 and $S_2 + A_2$ (S_2 and $S_1 + A_1$) in the state (21) is zero. Thus, the evolution of the properties that S_1 (S_2) has relative to $A_1 + A_2 + E$ is according to the dynamics of no entanglement, where e-nonlocality cannot occur. In more realistic models, the states of the environment are only approximately orthogonal relative to the pointer basis. This means that the degree of entanglement between S_1 and $S_2 + A_2$ (S_2 and $S_1 + A_1$) is virtually zero: The 'effective collapse' onto the pointer eigenstates induced by

decoherence reduces the degree of entanglement to approximately zero. Thus, given decoherence the universal dynamics of the properties of S_1 (S_2), as a subsystem of S_1+S_2, relative to A_1+A_2+E effectively reduces to the dynamics that these properties have in no entanglement. Accordingly, the probability of e-nonlocality in the value of R_1 (R_2) relative to $A_1 + A_2 + E$, e.g. the probability that these values will be different on the hypersurfaces σ_1 and σ_4 (σ_3) is virtually zero. More generally, based on these considerations, it is not difficult to show that the probability of e-nonlocality in experimental circumstances is virtually zero. (Note that while decoherence influences the dynamics of properties and accordingly plays a central role in accounting for our experience of the classical-like behavior of macroscopic systems, in contrast to some Everett-like interpretations, the decoherence-histories approach, Bub's modal interpretation and some other interpretations of quantum mechanics, it does not play any role in the property assignment.)

In fact, e-nonlocality would be unobservable even if the systems S_1, S_2, A_1, A_2 and observers of the pointer observables of A_1 and A_2, P_1 and P_2 respectively, were completely isolated from their environment. To see why, consider for example an observer O_1, who perceives the value of the pointer observable P_1 on the hypersurface σ_1 and compares this value with the value of P_1 on the hypersurface σ_4; where (as before) by 'the value of P_1 (O_1) on a hypersurface,' we mean the value that this observable has in the state that obtains on that hypersurface.[19] (To simplify terminology, in what follows O_1 will refer to both a physical system and an agent. Context will distinguish between these different uses.) Let B_1 be a brain observable associated with O_1's beliefs about the value of P_1, and let $M_{\sigma 1}$ be a brain observable associated with O_1's memory of the value that P_1 has on σ_1. Suppose that $S_1 + S_2 + A_1 + A_2 + O_1$ is completely isolated from the environment. Then, according to the universal dynamics, some (relational) values of P_1 (e.g. the value that P_1 has, as a property of $S_1 + A_1 + O_1$, relative to $A_2 + S_2$) may not be the same on σ_1 and σ_4. In order to observe such e-nonlocality, O_1 will have to reliably monitor and compare these values of P_1 on both σ_1 and σ_4. To monitor and remember the values of P_1 on σ_1, the value of B_1 and M_{σ_1} on σ_1 have to get correlated with the value of P_1 on σ_1. But, due to the lack of interaction with the environment, the relevant degree of entaglement will be substantive. Thus, if these values are correlated on σ_1, the universal dynamics dictates that they will also be correlated on σ_4. So O_1 will not be able to notice that the value of P_1 on σ_1 is different from its value on σ_4, and accordingly will not be able to observe the e-nonlocality in this value (for more details, see Berkovitz and Hemmo 2005a, pp. 392–4).

10 TOWARD A RELATIVISTIC MODAL INTERPRETATION

10.1 On the relativistic constraints

According to the standard understanding of special relativity, there is no foliation of spacetime into parallel spacelike hyperplanes that is preferred by the laws of physics: All foliations of spacetime are on equal footing. Moreover, special relativity also

requires that the descriptions of physical reality in different coordinate systems will be consistent with each other. In the special relativistic spacetime – the Minkowski spacetime – this requirement is satisfied when the descriptions of physical reality in different foliations of spacetime into parallel spacelike hyperplanes (and accordingly in different inertial frames of reference) are related to each other by the Lorentz transformations. Any adequate special-relativistic interpretation of quantum mechanics will have to satisfy these requirements and reproduce the empirical predictions of orthodox quantum mechanics (which has long been considered to be the basic standard of empirical adequacy).

In Myrvold's no-go theorem for relativistic modal interpretations, the constraints imposed by special relativity and quantum mechanics are expressed by the Relativistic Born Rule: The distribution of local properties of systems, i.e. properties that these systems have irrespective of the rest of the universe, on every spacelike hypersurface should be according to the Born Rule for arbitrary initial quantum state. The relational modal interpretation trivially satisfies the Relativistic Born Rule, as it postulates no such properties (see Section 8.1). All the properties assigned by this interpretation are relational and accordingly nonlocal: They are relations between distant systems. Yet, Myrvold's theorem may easily be modified to apply to relational quantities which have the same value on any two hypersurfaces intersecting the region in which the system they pertain to is located. Thus, the nonlocal nature of relational properties *per se* is insufficient for circumventing this theorem. The relational modal interpretation involves a more radical type of nonlocality, the so-called 'e-nonlocality': The dynamics postulated by this interpretation dictates that when the degree of entanglement between the relevant systems is nonzero, the value of an apparently local quantity may be different on spacelike hypersurfaces that intersect the region in which the system it pertains to is located. In the set up of Myrvold's theorem, where the relevant degrees of entanglement are significant, the probability that the value of e.g. R_1 relative to $A_1 + A_2$ is not the same on σ_1 and σ_4 (σ_3 and σ_4), is substantial (see Sections 7 and 8.1). Accordingly, Myrvold's theorem fails to apply to the relational modal interpretation.

In their theorem, Dickson and Clifton consider the spin properties that each of the particles in the EPR/B experiment has, and the dictates of special relativity are expressed by two conditions: Fundamental Lorentz-Invariance and Invariant Transition Probabilities. In this set up, the relational modal interpretation we outlined above (see Sections 5-8) satisfies these conditions. Yet, Fundamental Lorentz-Invariance and Invariant Transition Probabilities do not always hold. For example, in the set up of Myrvold's theorem the values of R_1 and R_2 (as properties of $S_1 + S_2$) relative to $A_1 + A_2$ are not Lorentz covariant and accordingly do not satisfy Fundamental Lorentz Invariance. For while the transition probabilities of these values are Lorentz covariant, due to the indeterministic dynamics the values themselves are not. In the next two subsections, we shall suggest two strategies for addressing this problem. The first strategy is to relate properties not only to other systems but also to spacelike hypersurfaces. And the second strategy is to regard the property assignment of

the relational modal interpretation as a source of information for Lorentz-covariant properties only under certain conditions.

10.2 Relativity and hypersurface dependence

One way to render the properties assigned by the relational modal interpretation Lorentz invariant is to postulate that properties of systems are not only relational to other systems but also to spacelike hypersurfaces. The idea is that for any partition of the universe into two distinct systems, α and β, the reduced state of α on any spacelike hypersurface σ prescribes the properties that the system α (and its subsystems) have relative to both the system β and the hypersurface σ. These properties are invariant across all inertial reference frames. For relativizing to spacelike hypersurfaces, properties are in essence properties of spacetime, and accordingly are by their very nature Lorentz invariant. Thus, although in the set up of Myrvold's theorem the value that R_1 has (as a property of $S_1 + S_2$) relative to $A_1 + A_2$ and the hypersurface σ_1 may be different from the value it has (as a property of $S_1 + S_2$) relative to $A_1 + A_2$ and the hypersurface σ_4, these relational values are not frame dependent: Relative to $A_1 + A_2$ and any spacelike hypersurface σ_i, R_1 has the same value in all inertial reference frames. Indeed, any family of parallel spacelike hyperplanes may be associated with an inertial reference frame, and thus it may be tempting to identify hyperplane dependence with frame dependence. But, there is a conceptual difference between frame-dependent and hyperplane-dependent properties: Properties that are hyperplane dependent may be frame independent (see Aharonov and Albert 1981, Fleming 1995 and Maudlin 1994 and 1996). Furthermore, in general hypersurface-dependent properties cannot be associated with certain inertial frames.

The arguments in Section 9 may easily be modified so as to show that this hypersurface dependence is unobservable in experimental circumstances, where due to environmentally-induced decoherence the degree of entanglement between the relevant systems is virtually zero. For it follows from the dynamics of the relational modal interpretation that in such circumstances, the probability that the apparent local properties of a system on any two spacelike hypersurfaces that intersect the region in which the system is located will not be the same, is virtually zero.[20]

Relativizing properties to spacelike hypersurfaces is in a sense a step in the direction of transforming quantum mechanics into a spacetime theory. Indeed, such properties are in need of explication. (For example, one may wonder what does it mean to be at position x relative to hypersurface σ_1 and at position y relative to σ_2?) Yet, given the universal dynamics of properties (see Section 7), the hypersurface-dependent properties postulated by the relational modal interpretation are related to classical-like, hypersurface-independent properties: When the (relevant) degree of entanglement is (virtually) zero, such hypersurface-dependent properties evolve like the corresponding classical-like, hypersurface-independent properties.

10.3 Relativity and entanglement

Another way to try to reconcile the relational modal interpretation with special relativity is to restrict the property assignment to circumstances in which the degree of entanglement between the relevant systems is zero. In these circumstances, the properties assigned by the relational interpretation appear to be Lorentz covariant. The problem with this approach is that, in general, environmentally-induced decoherence does not totally suppress entanglement. And as is not difficult to see from our discussion in Section 10.1, when the degree of entanglement is not zero the properties assigned by the relational modal interpretation (as presented in Sections 5-9) may not be Lorentz covariant, no matter how small the entanglement is.

This may suggest that the relationship between the quantum state of a system and the range of its possible properties should be less direct. The reasoning is as follows. In the relational modal interpretation, the dynamics of properties depends on the degree of entanglement between the relevant systems (which are determined by the relational properties and the transformations under consideration). The violation of Fundamental Lorentz Invariance can only occur when the degree of entanglement is nonzero, and the likelihood of such violation is proportional to the degree of entanglement. If the relevant degree of entanglement is approximately zero, the probability that the relational value of an observable of a system will not be the same on any two spacelike hypersurfaces that intersect the region in which the system is located, is virtually zero. That is, in such cases the probability that the properties assigned by the relational modal interpretation will violate Fundamental Lorentz Invariance is virtually zero. While it is inappropriate to assign to a system in a state that approaches a (relevant) zero degree of entanglement the properties that the relational modal interpretation assigns in state of no entanglement, it is plausible to have a very high degree of belief (which for all intents and purposes is indistinguishable from one) in the occurrence of these properties. In particular, while it is inappropriate to assign to systems in states of approximate decoherence the properties that the relational modal interpretation assigns in states of perfect decoherence, it is plausible to have a very high degree of belief in the occurrence of such properties. In short, here the idea is to adopt an epistemic interpretation wherein quantum-mechanical states provide information about objective properties of systems only under certain conditions. These conditions are fulfilled when the relevant degree of entanglement is virtually zero, as the case is when macroscopic systems undergo decoherence interactions with their environment.[21]

11 CONCLUSIONS

In this paper we proposed a new modal interpretation of quantum mechanics in terms of relational properties. We argued that this relational interpretation has several important merits. It offers a solution to the measurement problem that the mainstream modal interpretations encounter in non-ideal measurements and various decoherence circumstances, and explanation to the failure of the so-called ‘property composition’

and 'property decomposition' in these interpretations. Also, in contrast to all current modal interpretations, the relational modal interpretation does not postulate any preferred basis, and it circumvents all the no-go theorems for a relativistic modal interpretation. Furthermore, the relational interpretation seems to provide better prospects for developing a genuinely relativistic version of the modal interpretation.

It may be objected that the relational modal interpretation is quite radical. We do not find this objection compelling. Indeed, the picture of physical reality portrayed by this interpretation is very different from those portrayed by the non-relational interpretations of quantum mechanics. Yet, in the history of physics the conception of physical reality has undergone a number of radical changes. In fact, orthodox quantum mechanics and its mainstream interpretations themselves mark a radical shift from classical physics. We think that the merits of any interpretation of quantum mechanics have to be judged mainly on the basis of its consistency, its empirical adequacy, its explanatory power and its compatibility with other major theories of the physical realm. The relational modal interpretation seems to fare well on all these accounts. As far as we can see, it is consistent, it is empirically adequate, it is explanatory and it provides reasons to believe that quantum mechanics could be reconciled with special relativity. Further, a survey of other attempts to develop a relativistic interpretation of quantum mechanics may demonstrate that they similarly involve radical assumptions about the nature of physical reality.

Finally, we believe that our study may also be relevant for other major attempts to reconcile quantum mechanics with special relativity. But, this is a subject for a future study.

12 ACKNOWLEDGEMENTS

We are very grateful to the editors Bill Demopoulos and Itamar Pitowsky for giving us the opportunity to contribute to this volume. As students of Itamar, we feel like the intellectual grandsons of Jeff Bub, who was Itamar's doctoral advisor. Jeff's work has been an important source of influence on our thinking about quantum mechanics. In our investigation of the prospects of modal interpretations of quantum mechanics, we have learned and been inspired by Jeff's own work on this topic. Jeff suggested that a way to solve major problems that modal interpretations encounter in reproducing the classical-like behavior of macroscopic systems is to postulate that the properties of systems are given by preferred observables, which are distinguished from other observables in that their behavior is stable under decoherence interactions of macroscopic systems with their environment. In this paper, we have suggested that decoherence can play a major role in a different way. That is, we have outlined a new version of the modal interpretation wherein the dynamics of properties directly depends on the relevant degrees of entanglement, which in turn depend on environmentally-induced decoherence. And we have argued that this alternative way of thinking about the role of decoherence also provides better prospects for reconciling the modal interpretation with special relativity. JB would also like to thank Jeff and his wife Robin Schuster for their precious friendship and care, and for adopting

him into their family. For comments and discussions, we are very grateful to Itamar Pitowsky. For financial and other support, JB would like to thank the University of Maryland Baltimore County, the PPM group, Konstanz University and the Centre for Philosophy of Natural and Social Sciences, London School of Economics.

NOTES

[1] Here, in presenting the measurement problem, we followed the common assumption that perception of definite pointer reading is correlated with the value of some brain observable.

[2] Of course, in standard quantum mechanics there are observables in the Hilbert space of $S + M + O$ that possess definite values in states (2) and (3), but these observables do not correspond to definite pointer outcomes. Further, given the assumption that perception of definite pointer reading is correlated with the value of some brain observable, these observables do not correspond to any brain observables associated with experiences of such outcomes.

[3] See, for example, Zurek 1991, Giulini et al. 1996 and references therein.

[4] See, for example, Griffiths 1984 and Gell-Mann and Hartle 1993.

[5] For a review and analysis of the role of decoherence in quantum theory, see Bacciagaluppi 2005 and references therein.

[6] For reviews and analyses of modal interpretations, see Bacciagaluppi 1996, Dieks and Vermaas 1998, Dickson 2002 and references therein. In our review, we shall only focus on the property assignments of these interpretations. For as we shall see in Section 8, the particular details of their dynamics are irrelevant to the no-go theorems for relativistic modal interpretations. Some of these theorems make very general presuppositions about the dynamics, whereas other make no presuppositions at all.

[7] Bacciagaluppi, Donald and Vermaas (1995) propose alternatively that in degeneracy points the definite properties are the limits of the definite properties around these points.

[8] For the sake of simplicity and brevity, here and henceforth by P we shall denote both projections and properties that correspond to these projections. Context will distinguish between these different uses.

[9] Bene and Dieks (2002) propose a new perspectivalist version of the modal interpretation which also provides an explanation for this violation.

[10] One can show that our experience of the physical world may be accounted for even if there are various uncorrelated subsets of physical properties that are associated with our experience. For example, one can show that the experiences of different observers will be compatible with each other even if they are related to *different* contexts (see Berkovitz and Hemmo 2005a). But for want of space we shall not pursue this issue here.

[11] We thank Itamar Pitowsky for discussions of these measures.

[12] In Berkovitz and Hemmo (2006), we propose that the above dynamics can also be applied to modal interpretations wherein properties of composite systems are interpreted as holistic (non-relational) properties that are not decomposable into the properties of their subsystems.

[13] For a recent review of the implications of Bell's theorem for the nature of quantum non-locality, see Berkovitz (2006) and references therein.

[14] For a discussion of various types of non-locality that are compatible with relativity, see Maudlin 1994.

[15] For the sake of brevity, by 'Born rule' we mean a Born-like rule that applies to properties in general rather than only to measurement outcomes.

[16] Note that these values of R_1 and R_2 are not meant to be hypersurface-dependent properties, but rather the values that these observables have in the states that obtain on the hypersurfaces $\sigma_1 - \sigma_4$.

[17] In fact, Local Properties and the Relativistic Born Rule jointly imply Joint Probabilities. Yet, for the sake of clarifying the way the relational modal interpretation circumvents Myrvold's theorem, it is important to make Joint Probabilities an explicit assumption.

[18] Artnzenius (1998) analyzes Dickson and Clifton's argument. He proposes a theorem that relies on the assumption of the existence of joint probability distribution over the local possessed properties on the

spacelike hypersurfaces $\sigma_1 - \sigma_4$. In that respect, it is similar to Myrvold's theorem, which may be regarded as a generalization of it.

[19] The exact set up that would allow for the observations of these values of P_1 (O_1) need not concern us below. Here, all we need to assume is that such observations can be made.

[20] Again, note that in contrast to some other no-collapse interpretations, decoherence *per se* does not play any role in the property assignment.

[21] This is not to say that definite properties obtain only in states of environmentally-induced decoherence, as in Bub's modal interpretation, though such a view is consistent with the above reading of the relational modal interpretation. Rather, the idea here is that quantum-mechanical states are reliable sources of information about Lorentz covariant properties only when the relevant degrees of entanglement are virtually zero.

REFERENCES

Aharonov, Y. and Albert, D. (1981), 'Can we make sense out of the measurement process in relativistic quantum mechanics', *Physical Review D* **24**, 359–370.

Arntzenius, F. (1998), 'Curiouser and curiouser: problems for modal interpretations of quantum mechanics', in Dieks and Vermaas (eds.), pp. 337–377.

Bacciagaluppi, G. (1995), 'Kochen-Specker theorem in the modal interpretation of quantum mechanics', *International Journal of Theoretical Physics* **34**, 1206–1215.

Bacciagaluppi, G. (1996), *Topics in the modal interpretation of quantum mechanics*, Ph.D. Thesis, Cambridge University.

Bacciagaluppi, G. (2000), 'Delocalized properties in the modal interpretation of a continuous model of decoherence', *Foundations of Physics* **30**, 1431–1444.

Bacciagaluppi, G. (2005), 'The role of decoherence in quantum theory', in Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2005 Edition). http://plato.stanford.edu/archives/summer2005/entries/qm-decoherence/

Bacciagaluppi, G., Donald, M., and Vermaas, P. (1995), 'Continuity and discontinuity of definite properties in the modal interpretation', *Helvetica Physica Acta* **68**, 679–704.

Bacciagaluppi, G. and Hemmo, M. (1996), 'Modal interpretations, decoherence and measurements', *Studies in History and Philosophy of Modern Physics* **27**, 239–277.

Bene, G. and Dieks, D. (2002), 'A perspectival version of the modal interpretation of quantum mechanics and the origin of macroscopic behavior', *Foundations of Physics* **32**, 645–671.

Berkovitz, J. (2006), 'Action at a distance in quantum mechanics', forthcoming in Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*.

Berkovitz, J. and Hemmo, M. (2005a), 'Modal interpretations of quantum mechanics and relativity: a reconsideration', *Foundations of Physics* **35**, 373–397.

Berkovitz, J. and Hemmo, M. (2006), 'How to reconcile modal interpretations of quantum mechanics with relativity', *Philosophy of Science* (PSA2004), forthcoming.

Bohm, D. (1952), 'A suggested interpretation of the quantum theory in terms of "hidden variables" I, II', *Physical Review* **85**, 166–179 and 180–193.

Bub, J. (1992), 'Quantum mechanics without the projection postulate', *Foundations of Physics* **22**, 737–754.

Bub, J. (1997), *Interpreting the Quantum World* (Cambridge University Press, Cambridge).

Clifton, R. (1996), 'The properties of modal interpretations of quantum mechanics', *British Journal for the Philosophy of Science* **47**, 371–398.

Dickson, M. (2002), 'The modal interpretations of quantum theory', in Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2002 Edition). http://plato.stanford.edu/archives/win2002/entries/qm-modal/(2002)

Dickson M. and Clifton, R. (1998), 'Lorentz invariance in modal interpretations', in Dieks and Vermaas (eds.), pp. 9–47.

Dieks, D. (1989), 'Resolution of the measurement problem through decoherence of the quantum state', *Physics Letters A* **142**, 439–446.

Dieks, D. and Vermaas, P. (eds.) (1998), *The Modal Interpretation of Quantum Mechanics* (Kluwer, Dordrecht).

Everett, H. III (1957), 'Relative state' formulation of quantum mechanics', *Review of Modern Physics* **29**, 454–462.

Fleming, G. N. (1995), 'Just how radical is hyperplane dependence?', in R. Clifton (ed.) *Perspectives on Quantum Reality* (Kluwer, Dordrecht), pp. 11–28.

Fraassen, B. C. van (1991), *Quantum Mechanics: An Empiricist View* (Clarendon Press).

Gell-Mann, M. and Hartle. J. B. (1993), 'Classical equations for quantum systems', *Physical Review D* **47**, 3345–3382.

Ghirardi, G., Rimini, A., and Weber, T. (1986), 'Unified dynamics for microscopic and macroscopic systems', *Physical Review D* **34**, 470–479.

Ghirardi, G., Pearle, P., and Rimini, A. (1990), 'Markov processes in Hilbert space and continuous spontaneous localization of systems of identical particles', *Physical Review A* **42**, 78–89.

Giulini, D., Joos, E., Kiefer, C., Kupsch, J., Stamatescu, I., and Zeh, D. H. (1996), *Decoherence and the Appearance of a Classical World in Quantum Theory* (Springer, Berlin).

Griffiths, R. (1984), 'Consistent histories and the interpretation of quantum mechanics', *Journal of Statistical Physics* **36**, 219–272.

Healey, R. (1989), *The Philosophy of Quantum Mechanics: An Interactive Interpretation* (Cambridge University Press, Cambridge).

Joos, E. and Zeh, H. D. (1985), 'The emergence of classical properties through interaction with the environment', *Zeitschrift für Physik B* **59**, 223–243.

Kochen, S. (1985), 'A new interpretation of quantum mechanics', in P. Lahti and P. Mittelstaedt (eds.), *Symposium on the Foundations of Modern Physics* (World Scientific, Singapore), pp. 151–169.

Maudlin, T. (1994), *Quantum Nonlocality and Relativity* (Blackwell, Oxford).

Maudlin, T. (1996), 'Spacetime in the quantum world', in J. Cushing, A. Fine and S. Goldstein (eds.), *Bohmian Mechanics and Quantum Theory: An Appraisal* (Kluwer, Dordrecht), pp. 285–307.

Myrvold, W. (2002), 'Modal interpretations and relativity', *Foundations of Physics* **32**, 1773–1784.

Shimony, A. (1995), 'The degree of entanglement', in D. M. Greenberger and A. Zeilinger (eds.), *Fundamental Problems in Quantum Theory, Annals of the New York Academy of Sciences*, Volume **755**, pp. 675–679.

Vermaas, P. and Dieks, D. (1995), 'The modal interpretation of quantum mechanics and its generalization to density operators', *Foundations of Physics* **25**, 145–158.

Zurek, W. (1991), 'Decoherence and the transition from quantum to classical', *Physics Today* **44**, 36–44.

HARVEY R. BROWN* AND CHRISTOPHER G. TIMPSON†

2. WHY SPECIAL RELATIVITY SHOULD NOT BE A TEMPLATE FOR A FUNDAMENTAL REFORMULATION OF QUANTUM MECHANICS

The principle of relativity is a principle that narrows the possibilities; it is not a model, just as the second law of thermodynamics is not a model. Albert Einstein[1]

ABSTRACT

In a comparison of the principles of special relativity and of quantum mechanics, the former theory is marked by its relative economy and apparent explanatory simplicity. A number of theorists have thus been led to search for a small number of postulates—essentially information theoretic in nature—that would play the role in quantum mechanics that the relativity principle and the light postulate jointly play in Einstein's 1905 special relativity theory. The purpose of the present paper is to resist this idea, at least in so far as it is supposed to reveal the fundamental form of the theory. It is argued that the methodology of Einstein's 1905 theory represents a victory of pragmatism over explanatory depth, that its adoption only made sense in the context of the chaotic state state of physics at the start of the 20th century—as Einstein well knew.

1 QUANTUM MECHANICS: THE CBH THEOREM

In an important recent development in quantum mechanics, Clifton, Bub and Halvorson (henceforth CBH) have shown that the observables and state space of a physical theory must be quantum mechanical if three 'information-theoretic' constraints hold.[2] The constraints are:

1. no superluminal information transmission between two systems by measurement on one of them,
2. no broadcasting of information contained in an unknown physical state, and
3. no unconditionally secure bit-commitment.

The CBH theorem states that these constraints force any theory formulated in C^*-algebraic terms to incorporate a non-commuting algebra of observables for individual systems, kinematic independence for the algebras of space-like separated systems and the possibility of entanglement between space-like separated systems. (Conversely,

* Faculty of Philosophy, University of Oxford, 10 Merton Street, Oxford OX1 4JJ, U.K., harvey.brown@philosophy.ox.ac.uk

† Division of History and Philosophy of Science, School of Philosophy, University of Leeds, LS2 9JT, UK. c.g.timpson@leeds.ac.uk

W. Demopoulos and I. Pitowsky (eds.), Physical Theory and its Interpretation, 29–42.

any C^*-algebraic theory with these distinctively quantum properties will satisfy at least the three information-theoretic constraints.[3])

This result is not only of great interest in itself, but it appeared at a time when attention to the putatively fundamental role that the notion of information plays in understanding quantum theory has been growing significantly. It is not our aim in this paper to examine in detail either the scope of the theorem[4], or the contentious issue of the role of information in modern physics[5]. We are concerned with the methodological issues at stake. At the start of their paper, CBH wrote:

> The fact that one can characterize quantum theory ... in terms of just a few simple information-theoretic principles ... lends credence to the idea that an information-theoretic point of view is the right perspective to adopt in relation to quantum theory. Notice, in particular, that our derivation links information-theoretic principles directly to the very features of quantum theory—noncommutativity and nonlocality—that are so conceptually problematic from a purely physical/mechanical point of view. We therefore suggest substituting for the conceptually problematic mechanical perspective on quantum theory an information-theoretic perspective. That is, we are suggesting that quantum theory be viewed, not as first and foremost a mechanical theory of waves and particles ...but as a theory about the possibilities and impossibilities of information transfer.[6]

Even more significantly for our purposes, at the end of their paper CBH suggested an analogy between their characterization of quantum mechanics and Albert Einstein's special theory of relativity (henceforth SR). The "foundational significance" of the CBH derivation is, according to these authors, that quantum mechanics should be interpreted as a *principle theory*, in the sense of the term that Einstein used to describe his 1905 formulation of SR.[7] CBH saw their constraints as analogous to the principles—the relativity principle and the light postulate—used by Einstein to derive the nature of relativistic kinematics.

There can be no doubt that Einstein's 1905 treatment of relativistic kinematics was a triumph of economy in relation to the corresponding treatment of moving rods and (to the extent it existed, as we see below) clocks provided by the leading *fin de siècle* ether theorists. But it is still not sufficiently appreciated that by his own admission, Einstein's principle theory route was based on a policy of despair, and represented a strategic retreat from the more desirable but, in his view, temporarily unavailable *constructive* approach. It is worth dwelling a little on this historical episode, to see what implications it might have for the CBH program.[8]

2 SPECIAL RELATIVITY AS A "PRINCIPLE THEORY"

It is well known that the principle/constructive theory distinction was articulated by Einstein in a popular article on his theory of relativity published in 1919 in the London *Times*[9]. But it was a theme that appeared sporadically throughout his life-long writings.

In January 1908, roughly two and a half years after publishing his celebrated paper on special relativity[10], Einstein wrote in a letter to Arnold Sommerfeld:

> So, first to the question of whether I consider the relativistic treatment of, e.g., the mechanics of electrons as definitive. No, certainly not. It seems to me too that a physical theory can be satisfactory only when it builds up its structures from *elementary* foundations. The theory of relativity is not more conclusively and absolutely satisfactory than, for example, classical thermodynamics was before Boltzmann had interpreted entropy as probability. If the Michelson-Morley experiment had not put us in the worst predicament, no one would have perceived the relativity theory as a (half) salvation.[11]

Einstein is repeating here an analogy between SR and thermodynamics that he had mentioned in a published note addressed to Ehrenfest already in 1907, in which he compared SR with "the second law of the theory of heat."[12] In both cases, Einstein was emphasizing the *limitations* of SR, not its strengths.

In order to see why SR is only a 'half' salvation, consider for a minute the analogy with thermodynamics.

Think of an idealized single-piston heat engine undergoing a Carnot cycle, and consider the theoretical limits of its efficiency. Such limits can in principle be established by exploiting knowledge of the micro-structure of the working substance of the engine, and in particular by using the principles of statistical mechanics that apply to the molecular structure of the gas in the piston. A much easier approach, however, is to fall back on the laws of classical thermodynamics to shed light on the performance of the engine—phenomenological laws which stipulate nothing about the deep structure of the working substance. According to this approach, the efficiency of the heat engine must depend in a certain way on the ratio of the temperatures of the two heat reservoirs simply because, whatever the gas in the piston is made up of, if it did not it would be possible for the engine to act as a perpetual motion machine of 'the second kind'. And this possibility is simply ruled out by hypothesis in thermodynamics.

Yet it is hard to not to wonder why, after all, such a perpetual motion cannot exist. Indeed, it is widely held that statistical mechanics in principle explains why (even if the details involved are controversial). But thermodynamics cannot. The impossibility of perpetual motion machines of various kinds is the very starting point of thermodynamics. What this theory gains in practicality and in the evident empirical solidity of its premisses, it loses in providing physical insight.

Einstein considered thermodynamics as the archetypical example of what he would call in 1919 a principle theory in physics, one which is based on well verified, but unexplained observable regularities. On the other hand, statistical mechanics, or more specifically the kinetic theory of gases, was for Einstein the prime example of a constructive theory, one built on the "elementary foundations" mentioned in his 1908 letter. These foundations involve hypotheses about unseen fundamental processes—normally involving the microstructure of bodies and its mechanical principles.

The distinction has been the subject of increasing attention in recent years[13], but it is easily misunderstood. First, it is clearly not categorical: all theories have principles, it is just that some are more phenomenological than others. Thermodynamics and statistical mechanics are on opposite ends of a spectrum of possible theories, and there are indeed respectable theories—as we shall see below—which lie somewhere in between.

Principle theories are typically employed when constructive theories are either unavailable, too difficult to build, or relatively unwieldy. For according to Einstein, "when we say we have succeeded in understanding a group of natural processes, we invariably mean that a constructive theory has been found which covers the processes in question."[14] Yet, Einstein stressed that SR is a principle theory. Why then did he feel it necessary to sacrifice explanatory content in developing his theory of relativity?

3 RODS, CLOCKS, AND THE QUANTUM

Recall the title of Einstein's 1905 relativity paper: "On the electrodynamics of moving bodies". One of the great challenges of late nineteenth century electrodynamics and optics was to predict the outcome of experiments involving electromagnetic phenomena being performed in a laboratory *moving with respect to the luminiferous ether*. After all, the earth is in motion relative to the centre of mass of the solar system, and at least some of the time must be moving relative to the ether—the invisible seat of electromagnetic phenomena. But by the turn of the century, the ether had become in the minds of some experts a very shadowy entity indeed. Made of an obscure kind of "imponderable matter", its main role was increasingly just that of providing the inertial frame of reference relative to which the fundamental electromagnetic field equations of Maxwell were postulated to hold. The question was now: what form do the field equations have in earth-bound frames that are moving relative to this fundamental frame?

Einstein is famous for claiming in 1905, on the basis of his relativity principle, that all laws of physics, including those of electrodynamics, take the same form in all inertial reference frames, so happily Maxwell's equations can be used just as well in the moving laboratory frame. But this conclusion, or something very close to it, had already been anticipated by several great ether theorists, including H. A. Lorentz, Joseph Larmor and particularly Henri Poincaré. This was largely because there had been from the middle of the nineteenth century all the way to 1905 a series of experiments involving optical and electromagnetic effects that failed to show any sign of the ether wind rushing through the laboratory: it was indeed as if the earth was always at rest relative to the ether. (The most famous of these, and the most surprising, was of course the 1887 Michelson-Morley experiment.) Like the above-mentioned ether theorists, Einstein realized that the covariance of Maxwell's equations—the form invariance of the equations—is achieved when the relevant coordinate transformations take a very special form, but Einstein was unique in his understanding that these transformations, properly understood, encode new

predictions as to the behaviour of rigid bodies and clocks in motion. That is why, in Einstein's mind, a new understanding of space and time themselves was in the offing.

Both the mathematical form of the transformations, and at least the non-classical distortion of moving rigid bodies were already known to Lorentz, Larmor and Poincaré—indeed a family of possible deformation effects was originally suggested independently by Lorentz and G. F. FitzGerald to explain the Michelson-Morley result.[15] It was the connection between them, i.e. between the coordinate transformations and motion-induced deformation, that had not been fully appreciated before Einstein. In the first ("kinematical") part of his 1905 relativity paper, Einstein established the operational meaning of the so-called Lorentz coordinate transformations and showed that they lead not just to a special case of FitzGerald-Lorentz deformation (longitudinal contraction), but also to the "slowing down" of clocks in motion—the phenomenon of time dilation. Now it is still not well known that Larmor and Lorentz had come tantalizingly close to predicting this phenomenon; they had independently seen just before the turn of the century how it must hold in certain very special cases. But as a general effect that does not depend on the constitution of a clock, its discovery was Einstein's own.

Einstein did something else that was new and important in the kinematical part of his paper. He derived the Lorentz transformations not from the symmetry properties of Maxwell's equations, but by using an argument inspired by thermodynamics. The reason lies in his earlier investigations of the properties of black-body radiation.

Several months before he wrote his paper on SR, Einstein had written a revolutionary paper claiming that electromagnetic radiation has a granular structure. The suggestion that radiation was made of quanta—or photons as they would later be dubbed—was the basis of Einstein's extraordinary treatment of the photoelectric effect in the same paper. But the immediate consequence of Einstein's commitment to the photon was to destabilize in his mind all the previous work on the electrodynamics of moving bodies.

All the work of the ether theorists was based on the assumption that Maxwellian electrodynamics is strictly true, and not just true on average. In the work of Lorentz, Larmor and Poincaré, the Lorentz transformations make their appearance as symmetry transformations (whether considered approximate or otherwise) of these equations. But Maxwell's equations are incompatible with the existence of the photon.

In his 1949 *Autobiographical Notes*, published when he was 67, Einstein was clear about the seismic implications of this conundrum.

> Reflections of this type [on the dual wave-particle nature of radiation] made it clear to me as long ago as shortly after 1900, i.e., shortly after Planck's trailblazing work, that neither mechanics nor electrodynamics could (except in limiting cases) claim exact validity. By and by I despaired of the possibility of discovering the true laws by means of constructive efforts based on known facts.[16]

Already in the *Notes*, Einstein had pointed out that the general validity of Newtonian mechanics came to grief with the success of the electrodynamics of Faraday and

Maxwell, which led to Hertz's detection of electromagnetic waves—"phenomena which by their very nature are detached from every ponderable matter".[17] Later, he summarized the nature of Planck's 1900 derivation of his celebrated black-body radiation formula, in which quantization of absorption and emission of energy by the mechanical resonators is presupposed. Einstein noted that although this contradicted the received view, it was not immediately clear that electrodynamics—as opposed to mechanics—was violated. But now with the emergence of the light quantum, not even electrodynamics was sacrosanct.

> All my attempts . . . to adapt the theoretical foundation of physics to this [new type of] knowledge failed completely. It was if the ground had been pulled out from under one, with no firm foundation to be seen anywhere, upon which one could have built.[18]

Earlier in the *Notes*, Einstein had sung the praises of classical thermodynamics, "the only physical theory of universal content concerning which I am convinced that, within the framework of the applicability of its basic concepts, it will never be overthrown". Now, he explains how the very structure of the theory was influential in the search for a way out of the turn-of-the-century crisis in physics.

> The longer and more despairingly I tried, the more I came to the conviction that only the discovery of a universal formal principle could lead us to assured results. The example I saw before me was thermodynamics. The general principle was there given in the theorem[19]: the laws of nature are such that it is impossible to construct a *perpetuum mobile* (of the first and second kind). How, then, could such a universal principle be found?[20]

4 EINSTEIN'S DOUBTS

It is well-known that Einstein's based his derivation of the Lorentz transformations on a combination of the relativity principle (essentially the same as that defended by Newton) and his so-called light postulate. (The latter was the claim that relative to a certain inertial frame, the speed of light is independent of the speed of the source and isotropic—something every ether theorist took for granted when the frame in question is taken to be the fundamental ether rest frame[21] and something which remarkably Einstein felt would survive whatever the eventual quantum theory of radiation would reveal.) He showed that length contraction for rigid rods and time dilation for ideal clocks are consequences of these phenomenological assumptions, in the same way that, say, the existence of entropy and its non-decreasing behaviour over time for adiabatic systems are a consequence of the laws of thermodynamics. Of course, the precise form of the phenomena of contraction and dilation depended on Einstein's choice of a convention for spreading time through space in both the resting and moving frames—a choice Poincaré had already advocated.

Einstein would have preferred a constructive account of these relativistic effects, presumably based on the nature of the non-gravitational forces that hold the constituent parts of rods and clocks together. But as we have seen, for Einstein the elements

of such an account were not to be had in 1905. The price to be paid for the resulting strategic retreat to a principle theory approach was not just loss of insight; Einstein became increasingly uneasy about the role played by rods and clocks in this approach. This unease is seen in a paper entitled "Geometry and Experience" he published in 1921[22], and in particular in his 1949 *Autobiographical Notes*:

> One is struck [by the fact] that the theory [of special relativity] . . . introduces two kinds of physical things, i.e., (1) measuring rods and clocks, (2) all other things, e.g., the electromagnetic field, the material point, etc. This, in a certain sense, is inconsistent; strictly speaking measuring rods and clocks would have to be represented as solutions of the basic equations (objects consisting of moving atomic configurations), not, as it were, as theoretically self-sufficient entities. However, the procedure justifies itself because it was clear from the very beginning that the postulates of the theory are not strong enough to deduce from them sufficiently complete equations . . . in order to base upon such a foundation a theory of measuring rods and clocks. . . . But one must not legalize the mentioned sin so far as to imagine that intervals are physical entities of a special type, intrinsically different from other variables ('reducing physics to geometry', etc.).[23]

These remarks are noteworthy for several reasons.

First, there is the issue of justifying the "sin" of treating rods and clocks as primitive, or unstructured entities in SR. Einstein does not say in 1949, as he did in 1908 and 1921, that the "elementary" foundations of a constructive theory of matter are still unavailable; rather he simply reminds us of the limits built into the very form of the 1905 theory. It is hardly any justification at all. Considerable progress in the relativistic quantum theory of matter *had* been made between 1905 and 1949. Was it Einstein's long-standing distrust of the quantum theory that held him back from recognizing this progress and its implications for his formulation of SR?

Second, consider the criticism Abraham Pais made of H. A. Lorentz in his acclaimed 1982 biography of Einstein: "Lorentz never fully made the transition from the old dynamics to the new kinematics."[24] As late as 1915 Lorentz thought that the relativistic contraction of bodies in motion can be explained if the known property of distortion of the electrostatic field surrounding a moving charge is supposed to obtain for all the other forces that are involved in the cohesion of matter. In other words, Lorentz viewed such kinematical effects as length contraction as having a dynamical origin, and it is this notion that Pais found reprehensible. Yet, when Einstein appeals to the nature of rods and clocks as "moving atomic configurations", it seems that not even he ever fully accepted the distinction between dynamics and kinematics. For to say that length contraction is intrinsically kinematical would be like saying that energy or entropy are intrinsically thermodynamical, not mechanical—something Einstein would never have accepted.[25]

The limitations of Einstein's principle-theory approach to SR have been noted by a number of commentators since 1905, including Wolfgang Pauli and Arthur Eddington

in the 20s, W. F. G. Swann in the 40s, and Lajos Jánossy and John S. Bell in the 70s, and Dennis Dieks in the 80s.[26] All of these authors called for a more constructive version of SR. It was perhaps Bell who made the point in the clearest fashion.

> If you are, for example, quite convinced of the second law of thermodynamics, of the increase of entropy, there are many things that you can get directly from the second law which are very difficult to get directly from a detailed study of the kinetic theory of gases, but you have no excuse for not looking at the kinetic theory of gases to see how the increase of entropy actually comes about. In the same way, although Einstein's theory of special relativity would lead you to expect the FitzGerald contraction, you are not excused from seeing how the detailed dynamics of the system also leads to the FitzGerald contraction.[27]

What is remarkable is that Bell himself seemed to be unaware of Einstein's own distinction between principle and constructive theory, and his repeated references to the analogy between SR and thermodynamics. At any rate, Bell stressed that he had no "reservation whatever about the power and precision of Einstein's approach"; his main point was that "the longer road [a dynamical account of contraction and dilation] sometimes gives more familiarity with the country".[28]

5 THE CBH HISTORICAL FABLE

Let us return to the CBH argument. These authors offered a thought-provoking historical fable wherein SR began with Minkowski, who proposed a non-Newtonian geometry of space-time, and only later did Einstein come up with his principle theory approach. CHS regarded Minkowski as providing an "algorithm for relativistic kinematics", presumably based on the group of isometries of the postulated space-time structure, whereas in their fable they saw Einstein as furnishing an *interpretation* for SR: "a description of the conditions under which the [Minkowski] theory would be true, in terms of certain principles that constrain the law-like behaviour of physical systems". Analogously, it was argued, the CBH theorem could be viewed as providing an interpretation of quantum theory, based on information-theoretic constraints. It is clear from the CBH article that the authors regarded such an interpretation as having much in common with a position widely attributed to Niels Bohr, to the effect that quantum mechanics is not about micro-physical reality *per se* but rather the way we talk about it.

In attempting to evaluate CBH's neo-Bohrian stance, it is worth recalling first that the dominant viewpoint in the philosophy of space-time physics over the last few decades puts a very different gloss on Minkowski's contribution to SR. Far from being the basis of a mere algorithm for SR, the current orthodoxy seems to be that Minkowskian geometry provides a *constructive* dimension to SR (though it is not always put in these terms), and thereby significantly enhances its explanatory power. According to this view, it is the structure of the Minkowski space-time in which they are immersed that ultimately explains why rods and clocks in motion

contract and dilate respectively.[29] But it is also worth bearing in mind that this was not entirely Minkowski's own interpretation of the four-dimensional geometry that bears his name. Minkowski's original position was much more like Poincaré's (who indeed by 1906 had anticipated core features of Minkowski's work). It was that the Lorentz coordinate transformations can be seen as orthogonal transformations preserving the metrical properties of space-time, but the physical significance of these transformations derives from the fact that they are elements of the newly-discovered, or rather postulated, covariance group of all the non-gravitational interactions. The geometry does not come first—it is the dynamical symmetries that are fundamental, and susceptible to geometrical codification.[30] In short, on either of these two views of the significance of Minkowski's contribution, it amounts to a great deal more than a mere algorithm.

It is arguable that Minkowski's own reasoning is not at root incompatible with the currently unorthodox dynamical interpretation of relativistic kinematics outlined in the previous section. The starting point of this account is indeed the Lorentz covariance of the equations governing all the non-gravitational forces—which in turn account for the cohesive properties of rigid bodies and clocks. We are not dealing here with a fully-fledged constructive theory, because the full details of the quantum theory of such interactions (and quantum theory it must be) are not required in the story. But such a theory would go a long way to avoid Einstein's self-confessed "sin" of treating rods and clocks as structureless, primitive entities, and the treating of space-time intervals as entities of a special type in the explanatory scheme of things.

It is not our purpose here to defend this dynamical, semi-constructive approach to relativistic kinematics.[31] It is rather to point out that Einstein's original 1905 formulation of SR has its limitations, as Einstein himself knew full well and did not seek to hide. It is far from clear that he would have encouraged the use of SR—his 1905 SR—as a template for an 'interpretation' of quantum theory. Or rather, for a *fundamental* interpretation. It is a remarkable thing that what might be called the kinematic structure of quantum theory, the nature of its observables and state space structure, can it seems be given a principle-theory, or 'thermodynamic' underpinning. As Bell stressed, the beauty of thermodynamics is in its economy of reason, but the insight it provides is limited in relation to the messier story told in statistical mechanics.

In assessing the import of the CBH theorem, Jeffrey Bub wrote:

> Assuming the information-theoretic constrainsts are in fact satisfied in our world, no mechanical theory of quantum phenomena that includes an account of measurement interactions can be acceptable, and the appropriate aim of physics at the fundamental level becomes the representation and manipulation of information.[32]

The reasoning behind this remarkable conclusion that no mechanical account of the measurement process in quantum mechanics is viable, seems at first sight to be the analogue of the argument in SR that because Einstein treated rods and clocks as primitive entities in 1905, no analysis of their behaviour *qua* moving atomic configurations is appropriate. An argument flatly rejected by Einstein himself.

However, it should be noted that a key part of Bub's 2004 argument is that the historical success of statistical mechanics, and in particular recognition that the molecular-kinetic theory is more than a 'useful fiction', came about because of Einstein's theory of Brownian motion. This theory not only allowed molecules to be counted, but demonstrated the limits of validity of thermodynamics. Where, Bub effectively asks, is the analogue of such superiority of constructive thought—the analogue of fluctuation phenomena—in quantum mechanics?

> The methodological moral I draw from the thermodynamics case is simply that a mechanical theory that purports to solve the measurement problem is not acceptable if it can be shown that, *in principle*, the theory can have no excess empirical content over a quantum theory. By the CBH theorem, given the information-theoretic constraints any extension of a quantum theory, like Bohmian mechanics, must be empirically equivalent to a quantum theory, so no such theory can be acceptable as a deeper mechanical explanation of why quantum phenomena are such subject to the information-theoretic constraints. To be acceptable, a mechanical theory that includes an account of our measuring instruments as well as the quantum phenomena they reveal (and so purports to solve the measurement problem) *must violate one or more of the information-theoretic constraints*.[33]

Yet it is very doubtful whether Einstein advocated recognition of boosted rods and clocks as "moving atomic configurations" in SR because he thought such a view might ultimately lead to a violation of one or more of this 1905 postulates. It is more plausible that he did so because it made sense conceptually.[34] Likewise, disillusionment with the crude instrumentalistic nature of key aspects of Bohr's philosophy is justifiably one of the motivations for alternative interpretations of quantum theory—whether they involve an "extension" to the quantum formalism (such as the de Broglie-Bohm trajectories, or the collapse mechanism of GRW-type theories) or not (such as the Everett interpretation).[35]

ACKNOWLEDGMENTS

We wish to thank Bill Demopoulos for the kind invitation to contribute to this volume in honour of Jeff Bub, and to applaud him for conceiving and undertaking this project. We feel privileged. For nearly four decades Jeff Bub has been a leading figure in the foundations of quantum mechanics, through work characterized by honesty, rigour and penetration. Long may it continue.

NOTES

[1] This statement was made by Einstein in 1911 at a scientific meeting in Zurich; see Galison (2004), p. 268. In 1911 Einstein was still using "principle of relativity" to mean theory of relativity.

[2] Clifton et al. (2003).
[3] CBH showed (*op. cit*) that such quantum properties imply the first two constraints, and Halvorson (2004) showed that the third constraint related to bit-commitment also follows.
[4] In this connection see Valentini (2003) and Timpson (2004), section 9.2.2. We feel it worthwhile pointing out that in non-relativistic quantum mechanics, it has long been accepted that signalling at infinite speeds is a theoretical possibility. For example, a particle strictly confined to a region of compact support by means of a potential barrier can propagate to arbitrary distances in arbitrarily short times when the barrier is suddenly removed. This does not violate the no-signalling theorem in quantum mechanics because the latter is defined with respect to communication between pairs of entangled systems. But what this case emphasizes is that the no-superluminal-information-transmission constraint in the CBH theorem is of limited validity, at least in non-relativistic quantum mechanics.
[5] See Timpson (2002, 2003, 2004).
[6] Clifton et al. (2003), p. 4.
[7] Clifton et al. (2003), p. 24.
[8] The present paper, sections 2, 3 and 4 of which draw heavily on Brown (2005a, b), is a development of views expressed in Timpson (2004), section 9.2.
[9] Einstein (1919).
[10] Einstein (1905).
[11] Einstein (1995).
[12] Einstein (1907).
[13] See, for example, Brown and Pooley (2001, 2006) and Balashov and Janssen (2003).
[14] Einstein (1919).
[15] For recent treatments of this episode, see Brown (2001, 2005b).
[16] Einstein (1969), p. 51, 53.
[17] *Op. cit.*, p. 25.
[18] *Op. cit.*, p. 45.
[19] The word "theorem" for "Satze" in the translation by P. A. Schilpp is perhaps better rendered as "sentence" or "statement". One of us (H.R.B.) thanks Thomas Müller for discussion of this point.
[20] *Op. cit.*, p. 53.
[21] In 1921, Wolfgang Pauli would correctly describe Einstein's light postulate as the "true essence of the old aether point of view"; Pauli (1981), p. 5. It should also be noted that the derivation of the Lorentz transformations requires a third, admittedly innocuous, assumption: the isotropy of space.
[22] Einstein (1921).
[23] Einstein (1969), pp. 59, 61.
[24] Pais (1982), p. 167.
[25] Joseph Larmor commented in relation to Einstein's 1905 relativity paper that it actually contained dynamical reasoning "masquerading in the language of kinematics"; Larmor (1929), p. 644.
[26] See Pauli (1981); Eddington (1928), p. 7; Swann (1941); Jánossy (1971); Bell (1976, 1992); and Dieks (1984).
[27] Bell (1992).
[28] Bell (1976). For a discussion of Bell's 1976 treatment of SR by way of a "Lorentzian pedagogy", see Brown and Pooley (2001) and Brown (2005b).
[29] See Balashov and Janssen (2003) and Brown and Pooley (2006).
[30] See Brown (2005b), ch 8.
[31] For such a defense, see Brown and Pooley (2001, 2006) and Brown (2005b).
[32] Bub (2004), p. 242.
[33] Bub (2004), p. 261.
[34] It is however interesting to ask whether there actually is an analogue of Brownian motion in the dynamical interpretation of SR. A positive answer, which appeals to certain phenomena in quantum field theory such as the Scharnhorst effect, is defended in Brown (2005b), ch. 9.
[35] For further arguments in this vein, in particular defending the de Broglie-Bohm theory from Bub's 2004 criticism, see Timpson (2004), pp. 218–222.

REFERENCES

Yuri Balashov and Michel Janssen. Critical notice: Presentism and relativity. *British Journal for the Philosophy of Science*, 54:327–46, 2003.

John S. Bell. How to teach special relativity. *Progress in Scientific Culture*, 1, 1976. Reprinted in Bell (1987), pp. 67–80.

John S. Bell. *Speakable and Unspeakable in Quantum Mechanics*. Cambridge University Press, Cambridge, 1987.

John S. Bell. George Francis FitzGerald. *Physics World*, 5:31–35, 1992. 1989 lecture, abridged by Denis Weare.

Harvey R. Brown. The origins of length contraction: I the FitzGerald-Lorentz deformation hypothesis. *American Journal of Physics*, 69:1044–1054, 2001. E-prints: arXive:gr-qc/0104032; PITT-PHIL-SCI 218.

Harvey R. Brown. Einstein's misgivings about his 1905 formulation of special relativity. *European Journal of Physics*, 26:S85S90, special Einstein issue, 2005a.

Harvey R. Brown. *Physical relativity: space-time structure from a dynamical perspective*. Oxford University Press, Oxford, 2005b.

Harvey R. Brown and Oliver Pooley. The origins of the spacetime metric: Bell's Lorentzian pedagogy and its significance in general relativity. In Callender and Huggett (2001), pp. 256–272, 2001. E-print: arXive:gr-qc/9908048.

Harvey R. Brown and Oliver Pooley. Minkowski space-time: a glorious non-entity. To appear in *The ontology of spacetime*, Dennis Dieks (ed.), Elsevier, 2006. An earlier version appeared in arXive: physics/0403088 and PITT-PHIL-SCI 1661.

Jeffrey Bub. Why the quantum? *Studies in History and Philosophy of Modern Physics*, 35:241–66, 2004.

Craig Callender and Nick Huggett (ed.). *Physics meets philosophy at the Planck scale*. Cambridge University Press, Cambridge, 2001.

Rob Clifton, Jeffrey Bub, and Hans Halvorson. Characterizing quantum theory in terms of information theoretic constraints. *Foundations of Physics*, 33(11):1561, 2003. Page references to arXive: quant-ph/0211089.

Dennis Dieks. The "reality" of the Lorentz contraction. *Zeitschrift für allgemeine Wissenschaftstheorie*, 15:33–45, 1984.

Arthur S. Eddington. *The Nature of the Physical World*. Cambridge University Press, Cambridge, 1928.

A. Einstein. Zur elektrodynamik bewegter körper. *Annalen der Physik*, 17:891–921, 1905.

A. Einstein. Bemerkung zur Notiz des Herrn P. Ehrenfest: Translation deformierbarer Elektronen und der Flächensatz. *Annalen der Physik*, 23:206–208, 1907. English translation in Einstein (1989), Doc. 44, pp. 236–7.

A. Einstein. What is the theory of relativity? The London Times, 1919. Reprinted in Einstein (1982), pp. 227–32.

A. Einstein. Geometrie und erfahrung. *Erweite Fassung des Festvortrages gehalten an der presussischen Akademie* Springer: Berlin, 1921. Translated by S. Bargmann as 'Geometry and Experience' in Einstein (1982), pp. 232–246.

A. Einstein. Autobiographical notes. In P. A. Schilpp, editor, *Albert Einstein: Philosopher-Scientist, Vol. 1*, pages 1–94. Open Court, Illinois, 1969.

A. Einstein. Letter to Arnold Sommerfeld, January 14, 1908. Document 73 in *The Collected Papers of Albert Einstein, Vol. 5, The Swiss Years: Correspondence, 1902–1914 (English Translation Supplement*, M. J. Klein, A. J. Kox and R. Schulmann (eds.), Princeton University Press, Princeton. Translated by A. Beck., 1995.

Peter Galison. *Einstein's Clocks, Poincaré's Maps. Empires of Time*. Hodder and Stoughton, London, 2004.

Hans Halvorson. On information-theoretic characterizations of physical theories. *Studies in History and Philosophy of Modern Physics*, 35(2):277–93, 2004.

L. Jánossy. *Theory of Relativity based on Physical Reality*. Akadémie Kiadó, Budapest, 1971.

Joseph Larmor. *Mathematical and Physical Papers, Vol. I*. Cambridge University Press, Cambridge, 1929.

Abraham Pais. *'Subtle is the Lord ...' The Science and the Life of Albert Einstein*. Oxford University Press, New York, 1982.

Wolfgang Pauli. *Theory of Relativity*. Dover, New York, 1981. Originally published as 'Relativitätstheorie', *Encyklopädie der matematischen Wissenschaften, mit Einschluss ihrer Anwendungen vol 5. Physik* ed. A. Sommerfeld (Tauber, Leibzig), 1921; pp. 539–775.

W. F. G. Swann. Relativity, the FitzGerald-Lorentz contraction, and quantum theory. *Reviews of Modern Physics*, 13:197–202, 1941.

Christopher G. Timpson. The applicability of the Shannon information in quantum mechanics and Zeilinger's foundational principle. *Philosophy of Science, Supplement: Proceedings of PSA*, 70: 1233–44, 2002.

Christopher G. Timpson. On a supposed conceptual inadequacy of the Shannon information in quantum mechanics. *Studies in History and Philosophy of Modern Physics*, 33:441–68, 2003.

Christopher G. Timpson. Quantum information theory and the foundations of quantum mechanics. D.Phil. thesis, Oxford University, 2004.

Antony Valentini. Universal signature of non-quantum systems. arXive: quant-ph/0309107, 2003.

JEREMY BUTTERFIELD*

3. ON SYMMETRY AND CONSERVED QUANTITIES IN CLASSICAL MECHANICS

ABSTRACT

This paper expounds the relations between continuous symmetries and conserved quantities, i.e. Noether's "first theorem", in both the Lagrangian and Hamiltonian frameworks for classical mechanics. This illustrates one of mechanics' grand themes: exploiting a symmetry so as to reduce the number of variables needed to treat a problem.

For both frameworks, I emphasise that the theorem is underpinned by the idea of cyclic coordinates. In the Lagrangian framework, the main extra "ingredient" is the rectification of vector fields afforded by the local existence and uniqueness of solutions to ordinary differential equations. In the Hamiltonian framework, the main extra ingredients are the asymmetry of the Poisson bracket, and the fact that a vector field generates canonical transformations iff it is Hamiltonian.

1 INTRODUCTION

The strategy of simplifying a mechanical problem by exploiting a symmetry so as to reduce the number of variables is one of classical mechanics' grand themes. It is theoretically deep, practically important, and recurrent in the history of the subject. Indeed, it occurs already in 1687, in Newton's solution of the Kepler problem; (or more generally, the problem of two bodies exerting equal and opposite forces along the line between them). The symmetries are translations and rotations, and the corresponding conserved quantities are the linear and angular momenta.

This paper will expound one central aspect of this large subject. Namely, the relations between continuous symmetries and conserved quantities—in effect, Noether's "first theorem": which I expound in both the Lagrangian and Hamiltonian frameworks, though confining myself to finite-dimensional systems. As we shall see, this

* All Souls College, Oxford OX1 4AL, email: jb56@cus.cam.ac.uk. It is a pleasure to dedicate this paper to Jeff Bub, who has made such profound contributions to the philosophy of quantum theory. Though the paper is about classical, not quantum, mechanics, I hope that with his love of geometry, he enjoys symplectic forms as much as inner products!

W. Demopoulos and I. Pitowsky (eds.), Physical Theory and its Interpretation, 43–100.

topic is underpinned by the theorems in elementary Lagrangian and Hamiltonian mechanics about cyclic (ignorable) coordinates and their corresponding conserved momenta. (Again, there is a glorious history: these theorems were of course clear to these subjects' founders.) Broadly speaking, my discussion will make increasing use, as it proceeds, of the language of modern geometry. It will also emphasise Hamiltonian, rather than Lagrangian, mechanics: apart from mention of the Legendre transformation, the Lagrangian framework drops out wholly after Section 3.4.1.[1]

There are several motivations for studying this topic. As regards physics, many of the ideas and results can be generalized to infinite-dimensional classical systems; and in either the original or the generalized form, they underpin developments in quantum theories. The topic also leads into another important subject, the modern theory of symplectic reduction: (for a philosopher's introduction, cf. Butterfield (2006)). As regards philosophy, the topic is a central focus for the discussion of symmetry, which is both a long-established philosophical field and a currently active one: cf. Brading and Castellani (2003). (Some of the current interest relates to symplectic reduction, whose philosophical significance has been stressed recently, especially by Belot: Butterfield (2006) gives references.)

The plan of the paper is as follows. In Section 2, I review the elements of the Lagrangian framework, emphasising the elementary theorem that cyclic coordinates yield conserved momenta, and introducing the modern geometric language in which mechanics is often cast. Then I review Noether's theorem in the Lagrangian framework (Section 3). I emphasise how the theorem depends on two others: the elementary theorem about cyclic coordinates, and the local existence and uniqueness of solutions of ordinary differential equations. Then I introduce Hamiltonian mechanics, again emphasising how cyclic coordinates yield conserved momenta; and approaching canonical transformations through the symplectic form (Section 4). This leads to Section 5's discussion of Poisson brackets; and thereby, of the Hamiltonian version of Noether's theorem. In particular, we see what it would take to prove that this version is more powerful than (encompasses) the Lagrangian version. By the end of the Section, it only remains to show that a vector field generates a one-parameter family of canonical transformations iff it is a Hamiltonian vector field. It turns out that we can show this without having to develop much of the theory of canonical transformations. We do so in the course of the final Section's account of the geometric structure of Hamiltonian mechanics, especially the symplectic structure of a cotangent bundle (Section 6). Finally, we end the paper by mentioning a generalized framework for Hamiltonian mechanics which is crucial for symplectic reduction. This framework takes the Poisson bracket, rather than the symplectic form, as the basic notion; with the result that the state-space is, instead of a cotangent bundle, a generalization called a 'Poisson manifold'.

2 LAGRANGIAN MECHANICS

2.1 Lagrange's equations

We consider a mechanical system with n configurational degrees of freedom (for short: *n freedoms*), described by the usual *Lagrange's equations*. These are n second-order ordinary differential equations:

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}^i}\right) - \frac{\partial L}{\partial q^i} = 0, \quad i = 1, \ldots, n; \tag{2.1}$$

where the Lagrangian L is the difference of the kinetic and potential energies: $L := K - V$. (We use K for the kinetic energy, not the traditional T; for in differential geometry, we will use T a lot, both for 'tangent space' and 'derivative map'.)

I should emphasise at the outset that several special assumptions are needed in order to deduce eq. 2.1 from Newton's second law, as applied to the system's component parts: (assumptions that tend to get forgotten in the geometric formulations that will dominate later Sections!) But I will not go into many details about this, since:

(i) there is no single set of assumptions of minimum logical strength (nor a single "best package-deal" combining simplicity and minimum logical strength);
(ii) full discussions are available in many textbooks (or, from a philosophical viewpoint, in Butterfield 2004a: Section 3).

I will just indicate a simple and commonly used sufficient set of assumptions. But owing to (i) and (ii), the details here will not be cited in later Sections.

Note first that if the system consists of N point-particles (or bodies small enough to be treated as point-particles), so that a configuration is fixed by $3N$ cartesian coordinates, we may yet have $n < 3N$. For the system may be subject to constraints and we will require the q^i to be independently variable. More specifically, let us assume that any constraints on the system are *holonomic*; i.e. each is expressible as an equation $f(r^1, \ldots, r^m) = 0$ among the coordinates r^k of the system's component parts; (here the r^k could be the $3N$ cartesian coordinates of N point-particles, in which case $m := 3N$). A set of c such constraints can in principle be solved, defining a $(m - c)$-dimensional hypersurface Q in the m-dimensional space of the rs; so that on the *configuration space* Q we can define $n := m - c$ independent coordinates $q^i, i = 1, \ldots, n$.

Let us also assume that any constraints on the system are: (i) *scleronomous*, i.e. independent of time, so that Q is identified once and for all; (ii) *ideal*, i.e. the forces that maintain the constraints would do no work in any possible displacement consistent with the constraints and applied forces (a 'virtual displacement'). Let us also assume that the forces applied to the system are *monogenic*: i.e. the total work δw done in an infinitesimal virtual displacement is integrable; its integral is the *work function* U. (The term 'monogenic' is due to Lanczos (1986, p. 30), but followed by others e.g. Goldstein et al. (2002, p. 34).) And let us assume that the system is *conservative*: i.e. the work function U is independent of both the time and the generalized velocities $\dot{q}_i$, and depends only on the q^i: $U = U(q^1, \ldots, q^n)$.

So to sum up: let us assume that the constraints are holonomic, scleronomous and ideal, and that the system is monogenic with a velocity-independent work-function. Now let us define K to be the kinetic energy; i.e. in cartesian coordinates, with k now labelling particles, $K := \Sigma_k \frac{1}{2} m_k \mathbf{v}_k^2$. Let us also define $V := -U$ to be the potential energy, and set $L := K - V$. Then the above assumptions imply eq. 2.1.[2]

To solve mechanical problems, we need to integrate Lagrange's equations. Recall the idea from elementary calculus that n second-order ordinary differential equations have a (locally) unique solution, once we are given $2n$ arbitrary constants.Broadly speaking, this idea holds good for Lagrange's equations; and the $2n$ arbitrary constants can be given just as one would expect—as the initial configuration and generalized velocities $q^i(t_0), \dot{q}^i(t_0)$ at time t_0. More precisely: expanding the time derivatives in eq. 2.1, we get

$$\frac{\partial^2 L}{\partial \dot{q}^j \partial \dot{q}^i} \ddot{q}^j = -\frac{\partial^2 L}{\partial q^j \partial \dot{q}^i} \dot{q}^j - \frac{\partial^2 L}{\partial t \partial \dot{q}^i} + \frac{\partial L}{\partial \dot{q}^i} \tag{2.2}$$

so that the condition for being able to solve these equations to find the accelerations at some initial time t_0, $\ddot{q}^i(t_0)$, in terms of $q^i(t_0), \dot{q}^i(t_0)$ is that the *Hessian* matrix $\frac{\partial^2 L}{\partial \dot{q}^i \partial \dot{q}^j}$ be nonsingular. Writing the determinant as $|\ \ |$, and partial derivatives as subscripts, the condition is that:

$$\left| \frac{\partial^2 L}{\partial \dot{q}^j \partial \dot{q}^i} \right| \quad \equiv \quad | L_{\dot{q}^j \dot{q}^i} | \quad \neq \quad 0. \tag{2.3}$$

This *Hessian condition* holds in very many mechanical problems; and henceforth, we assume it. (If it fails, we enter the territory of constrained dynamics; for which cf. e.g. Henneaux and Teitelboim (1992, Chapters 1–5).) It underpins most of what follows: for it is needed to define the Legendre transformation, by which we pass from Lagrangian to Hamiltonian mechanics.

Of course, even with eq. 2.3, it is still in general hard *in practice* to solve for the $\ddot{q}^i(t_0)$: they are buried in the lhs of eq. 2.2. In (5) of Section 2.2.2, this will motivate the move to Hamiltonian mechanics.[3]

Given eq. 2.3, and so the accelerations at the initial time t_0, the basic theorem on the (local) existence and uniqueness of solutions of ordinary differential equations can be applied. (We will state this theorem in Section 3.4 in connection with Noether's theorem.)

By way of indicating the rich theory that can be built from eq. 2.1 and 2.3, I mention one main aspect: the power of variational formulations. Eq. 2.1 are the Euler-Lagrange equations for the variational problem $\delta \int L \ dt = 0$; i.e. they are necessary and sufficient for the action integral $I = \int L \ dt$ to be stationary. But variational principles will play no further role in this paper; (Butterfield 2004 is a philosophical discussion).

But our main concern, here and throughout this paper, is how symmetries yield conserved quantities, and thereby reduce the number of variables that need to be

considered in solving a problem. In fact, we are already in a position to prove Noether's theorem, to the effect that any (continuous) symmetry of the Lagrangian L yields a conserved quantity. But we postpone this to Section 3, until we have developed some more notions, especially geometric ones.

We begin with the idea of generalized momenta, and the result that the generalized momentum of any cyclic coordinate is a constant of the motion: though very simple, this result is the basis of Noether's theorem. Elementary examples prompt the definition of the *generalized*, or *canonical*, *momentum*, p_i, *conjugate* to a coordinate q^i as: $\frac{\partial L}{\partial \dot{q}^i}$; (this was first done by Poisson in 1809). Note that p_i need not have the dimensions of momentum: it will not if q^i does not have the dimension length. So Lagrange's equations can be written:

$$\frac{d}{dt}p_i = \frac{\partial L}{\partial q^i}; \tag{2.4}$$

We say a coordinate q^i is *cyclic* if L does not depend on q^i. (The term comes from the example of an angular coordinate of a particle subject to a central force. Another term is: *ignorable*.) Then the Lagrange equation for a cyclic coordinate, q^n say,becomes $\dot{p}_n = 0$, implying

$$p_n = \text{constant, } c_n \text{ say.} \tag{2.5}$$

So: the generalized momentum conjugate to a cyclic coordinate is a constant of the motion.

It is straightforward to show that this simple result encompasses the elementary theorems of the conservation of momentum, angular momentum and energy: this last corresponding to time's being a cyclic coordinate. As a simple example, consider the angular momentum of a free particle. The Lagrangian is, in spherical polar coordinates,

$$L = \frac{1}{2}m\left(\dot{r}^2 + r^2\dot{\theta}^2 + r^2\dot{\phi}^2\sin^2\theta\right) \tag{2.6}$$

so that $\partial L/\partial\phi = 0$. So the conjugate momentum

$$\frac{\partial L}{\partial \dot{\phi}} = mr^2\dot{\phi}\sin^2\theta, \tag{2.7}$$

which is the angular momentum about the z-axis, is conserved.

2.2 Geometrical perspective

2.2.1 Some restrictions of scope I turn to give a brief description of the elements of Lagrangian mechanics in terms of modern differential geometry. Here 'brief'

indicates that:

(i) I will assume without explanation various geometric notions, in particular: manifold, vector, 1-form (covector), metric, Lie derivative and tangent bundle.

(ii) I will disregard issues about degrees of smoothness: all manifolds, scalars, vectors etc. will be assumed to be as smooth as needed for the context.

(iii) I will also simplify by speaking "globally, not locally". I will speak as if the scalars, vector fields etc. are defined on a whole manifold; when in fact all that we can claim in application to most systems is a corresponding local statement—because for example, differential equations are guaranteed the existence and uniqueness only of a *local* solution.[4]

We begin by assuming that the configuration space (i.e. the constraint surface) Q is a manifold. The physical state of the system, taken as a pair of configuration and generalized velocities, is represented by a point in the tangent bundle TQ (also known as 'velocity phase space'). That is, writing T_x for the tangent space at $x \in Q$, TQ has points $(x, \tau), x \in Q, \tau \in T_x$. We will of course often work with the natural coordinate systems on TQ induced by coordinate systems q on Q; i.e. with the $2n$ coordinates $(q, \dot{q}) \equiv (q^i, \dot{q}^i)$.

The main idea of the geometric perspective is that this tangent bundle is the arena for Lagrangian mechanics. So various previous notions and results are now expressed in terms of the tangent bundle. In particular, the Lagrangian is a scalar function $L : TQ \to \mathbb{R}$ which "determines everything". And the conservation of the generalized momentum p_n conjugate to a cyclic coordinate q_n, $p_n \equiv p_n(q, \dot{q}) = c_n$, means that the motion of the system is confined to a level set $p_n^{-1}(c_n)$: where this level set is a $(2n - 1)$-dimensional sub-manifold of TQ.

But I must admit at the outset that working with TQ involves limiting our discussion to (a) time-independent Lagrangians and (b) time-independent coordinate transformations.

(a) Recall Section 2.1's assumptions that secured eq. 2.1. Velocity-dependent potentials and-or rheonomous constraints would prompt one to use what is often called the'extended configuration space' $Q \times \mathbb{R}$, and-or the 'extended velocity phase space' $TQ \times \mathbb{R}$.

(b) So would time-dependent coordinate transformations. This is a considerable limitation from a philosophical viewpoint, since it excludes boosts, which are central to the philosophical discussion of spacetime symmetry groups, and especially of relativity principles. To give the simplest example: the Lagrangian of a free particle is just its kinetic energy, which can be made zero by transforming to the particle's rest frame; i.e. it is not invariant under boosts.

2.2.2 The tangent bundle With these limitations admitted, we now describe Lagrangian mechanics on TQ, in five extended comments.

(1) $2n$ *first-order equations; the Hessian again*

The Lagrangian equations of motion are now $2n$ *first-order* equations for the functions $q^i(t), \dot{q}^i(t)$, falling in to two groups:

(a) the n equations eq. 2.2, with the $\ddot{q}^i$ taken as the time derivatives of $\dot{q}^i$ with respect to t; i.e. we envisage using the Hessian condition eq. 2.3 to solve eq. 2.2 for the $\ddot{q}^i$, hard though this usually is to do in practice;

(b) the n equations $\dot{q}^i = \frac{dq^i}{dt}$.

(2) *Vector fields and solutions*

(a) These $2n$ first-order equations are equivalent to a vector field on TQ: the 'dynamical vector field', or for short the 'dynamics'. I write it as D (to distinguish it from the generic vector field $X, Y, \ldots$).

(b) In the natural coordinates $(q^i, \dot{q}^i)$, the vector field D is expressed as

$$D = \dot{q}^i \frac{\partial}{\partial q^i} + \ddot{q}^i \frac{\partial}{\partial \dot{q}^i}; \tag{2.8}$$

and the rate of change of any dynamical variable f, taken as a scalar function on $TQ, f(q, \dot{q}) \in \mathbb{R}$, is given by

$$\frac{df}{dt} = \dot{q}^i \frac{\partial f}{\partial q^i} + \ddot{q}^i \frac{\partial f}{\partial \dot{q}^i} = D(f). \tag{2.9}$$

(c) So the Lagrangian L determines the dynamical vector field D, and so (for given initial $q, \dot{q}$) a (locally unique) solution: an integral curve of D, $2n$ functions of time $q(t), \dot{q}(t)$ (with the first n functions determining the latter). This separation of solutions/trajectories within TQ is important for the visual and qualitative understanding of solutions.

(3) *Canonical momenta are 1-forms*

Any point transformation, or any coordinate transformation $(q^i) \to (q'^i)$, in the configuration manifold Q, induces a basis-change in the tangent space T_q at $q \in Q$. Consider any vector $\tau \in T_q$ with components $\dot{q}^i$ in coordinate system (q^i) on Q, i.e. $\tau = \frac{d}{dt} = \dot{q}^i \frac{\partial}{\partial q^i}$; (think of a motion through configuration q with generalized velocity τ). Its components $\dot{q}'^i$ in the coordinate system (q'^i) (i.e. $\tau = \dot{q}'^i \frac{\partial}{\partial q'^i}$) are given by applying the chain rule to $q'^i = q'^i(q^k)$:

$$\dot{q}'^i \equiv \frac{\partial q'^i}{\partial q^k} \dot{q}^k. \tag{2.10}$$

so that we can "drop the dots":

$$\frac{\partial \dot{q}'^i}{\partial \dot{q}^j} = \frac{\partial q'^i}{\partial q^j}. \tag{2.11}$$

One easily checks, using eq. 2.11, that for any L, the canonical momenta $p_i := \frac{\partial L}{\partial \dot{q}^i}$ form a 1-form on Q, transforming under $(q^i) \to (q'^i)$ by:

$$p'_i := \frac{\partial L'}{\partial \dot{q}'^i} = \frac{\partial q^k}{\partial q'^i} \frac{\partial L}{\partial \dot{q}^k} \equiv \frac{\partial q^k}{\partial q'^i} p_k \tag{2.12}$$

That is, the canonical momenta defined by L form a 1-form field on Q. (We will later describe this as a cross-section of the cotangent bundle.)

(4) *Geometric formulation of Lagrange's equations*
We can formulate Lagrange's equations in a coordinate-independent way, by using three ingredients, namely:

(i) L itself (a scalar, so coordinate-independent);
(ii) the vector field D that L defines; and
(iii) the 1-form on TQ defined locally, in terms of the natural coordinates $(q^i, \dot{q}^i)$, by

$$\theta_L := \frac{\partial L}{\partial \dot{q}^i} dq^i. \tag{2.13}$$

(So the coefficients of θ_L for the other n elements of the dual basis, the $d\dot{q}^i$, are defined to be zero.) This 1-form is called the *canonical 1-form*. We shall see that it plays a role in Noether's theorem, and is centre-stage in Hamiltonian mechanics.

We combine these three ingredients using the idea of the Lie derivative of a 1-form along a vector field.

We will write the Lie derivative of θ_L along the vector field D on TQ, as $\mathcal{L}_D\theta_L$. (It is sometimes written as L; but we need the symbol L for the Lagrangian—and later on, for left translation.) By the Leibniz rule, $\mathcal{L}_D\theta_L$ is:

$$\mathcal{L}_D\theta_L = \left(\mathcal{L}_D \frac{\partial L}{\partial \dot{q}^i}\right) dq^i + \frac{\partial L}{\partial \dot{q}^i} \mathcal{L}_D(dq^i). \tag{2.14}$$

But the Lie derivative of any scalar function $f : TQ \to \mathbb{R}$ along any vector field X is just $X(f)$; and for the dynamical vector field D, this is just $\dot{f} = \frac{\partial f}{\partial q^i}\dot{q}^i + \frac{\partial f}{\partial \dot{q}^i}\ddot{q}^i$. So we have

$$\mathcal{L}_D\theta_L = \left(\frac{d}{dt}\frac{\partial L}{\partial \dot{q}^i}\right) dq^i + \frac{\partial L}{\partial \dot{q}^i} d\dot{q}^i. \tag{2.15}$$

Rewriting the first term by the Lagrange equations, we get

$$\mathcal{L}_D\theta_L = \left(\frac{\partial L}{\partial q^i}\right) dq^i + \frac{\partial L}{\partial \dot{q}^i} d\dot{q}^i \equiv dL. \tag{2.16}$$

We can conversely deduce the familiar Lagrange equations from eq. 2.16, by taking coordinates. So we conclude that these equations' coordinate-independent form is:

$$\mathcal{L}_D\theta_L = dL. \tag{2.17}$$

(5) *Towards the Hamiltonian framework*
Finally, a comment about the Lagrangian framework's limitations as regards solving problems, and how they prompt the transition to Hamiltonian mechanics.

Recall the remark at the end of Section 2.1, that the n equations eq. 2.2 are in general hard to solve for the $\ddot{q}^i(t_0)$: they lie buried in the left hand side of eq. 2.2. On the other hand, the n equations $\dot{q}^i = \frac{dq^i}{dt}$ (the second group of n equations in (1) above) are as simple as can be.

This makes it natural to seek another $2n$-dimensional space of variables, ξ^α say ($\alpha = 1, \ldots, 2n$), in which:

(i) a motion is described by first-order equations, so that we have the same advantage as in TQ that a unique trajectory passes through each point of the space; but in which

(ii) all $2n$ equations have the simple form $\frac{d\xi^\alpha}{dt} = f_\alpha(\xi^1, \ldots, \xi^{2n})$ for some set of functions $f_\alpha (\alpha = 1, \ldots, 2n)$.

Indeed, Hamiltonian mechanics provides exactly such a space: it is usually the cotangent bundle of the configuration manifold, instead of its tangent bundle. But before turning to that, we expound Noether's theorem in the current Lagrangian framework.

3 NOETHER'S THEOREM IN LAGRANGIAN MECHANICS

3.1 Preamble: a modest plan

Any discussion of symmetry in Lagrangian mechanics must include a treatment of "Noether's theorem". The scare quotes are to indicate that there is more than one Noether's theorem. Quite apart from Noether's work in other branches of mathematics, her paper (1918) on symmetries and conserved quantities in Lagrangian theories has several theorems. I will be concerned *only* with applying her first theorem to finite-dimensional systems. In short: it provides, for any continuous symmetry of a system's Lagrangian, a conserved quantity called the 'momentum conjugate to the symmetry'.

I stress at the outset that the great majority of subsequent applications and commentaries (also for her other theorems, besides her first) are concerned with versions of the theorems for infinite (i.e. continuous) systems. In fact, the context of Noether's investigation was contemporary debate about how to understand conservation principles and symmetries in the "ultimate classical continuous system", viz. gravitating matter as described by Einstein's general relativity. This theory can be given a Lagrangian formulation: that is, the equations of motion, i.e. Einstein's field equations, can be deduced from a Hamilton's Principle with an appropriate Lagrangian. The contemporary debate was especially about the conservation of energy and the principle of general covariance (also known as: diffeomorphism invariance). General covariance prompts one to consider how a variational principle transforms under spacetime coordinate transformations that are arbitrary, in the sense of varying from point to point. This leads to the idea of "local" symmetries, which since Noether's time has been immensely fruitful in both classical and quantum physics, and in both a Lagrangian and Hamiltonian framework.[5]

So I agree that from the perspective of Noether's work, and its enormous later development, this Section's application of the first theorem to finite-dimensional systems is, as they say, "trivial". Furthermore, this application is easily understood, *without* having to adopt that perspective, or even having to consider infinite systems. In other words: its statement and proof are natural, and simple, enough that the nineteenth century masters of mechanics, like Hamilton, Jacobi and Poincaré, would certainly recognize it in their own work—allowing of course for adjustments to modern language. In fact, versions of it for the Galilei group of Newtonian mechanics and the Lorentz group of special relativity were published a few years before Noether's paper; (Brading and Brown (2003, p. 90); for details, cf. Kastrup (1987)).[6]

Nevertheless, it is worth expounding the finite-system version of Noether's first theorem. For:

(i) It generalizes Section 2.1's result about cyclic coordinates, and thereby the elementary theorems of the conservation of momentum, angular momentum and energy which that result encompasses. The main generalization is that the theorem does not assume we have identified a cyclic coordinate. But on the other hand: every symmetry in the Noether sense will arise from a cyclic coordinate in some system q of generalized coordinates. (As we will see, this follows from the local existence and uniqueness of solutions of ordinary differential equations.)

(ii) This exposition will also prepare the way for our discussion of symmetry and conserved quantities in Hamiltonian mechanics.[7]

In this exposition, I will also discuss *en passant* the distinction between:

(i) the notion of symmetry at work in Noether's theorem, i.e. a symmetry of L, often called a *variational symmetry*; and

(ii) the notion of a symmetry of the set of solutions of a differential equation: often called a *dynamical symmetry*. This notion applies to all sorts of differential equations, and systems of them; not just to those with the form of Lagrange's equations (i.e. derivable from an variational principle). In short, this sort of symmetry is a map that sends any solution of the given equation(s) (in effect: a dynamically possible history of the system—a curve in the state-space) to some other solution. Finding such symmetries, and groups of them, is a central part of the modern theory of integration of differential equations (both ordinary and partial).

Broadly speaking, this notion is more general than that of a symmetry of L. Not only does it apply to many other sorts of differential equation. Also, for Lagrange's equations: a symmetry of L is (with one *caveat*) a symmetry of the solutions, i.e. a dynamical symmetry—but the converse is false.[8]

In this Section, the plan is as follows. We define:

(i) a *(continuous) symmetry* as a vector field (on the configuration manifold Q) that generates a family of transformations under which the Lagrangian is invariant;(Section 3.2);

(ii) the *momentum conjugate to a vector field*, as (roughly) the rate of change of the Lagrangian with respect to the $\dot{q}$s in the direction of the vector field; (Section 3.3).

These two definitions lead directly to Noether's theorem (Section 3.4): after all the stage-setting, the proof will be a one-liner application of Lagrange's equations.

3.2 Vector fields and symmetries—variational and dynamical

I need to expound three topics:
(1) the idea of a vector field on the configuration manifold Q; and how to lift it to TQ;
(2) the definition of a variational symmetry;
(3) the contrast between (2) and the idea of dynamical symmetry.

Note that, as in previous Sections, I will often speak, for simplicity, "globally, not locally", i.e. as if the relevant scalar functions, vector fields etc. are defined on all of Q or TQ. Of course, they need not be.

3.2.1 Vector fields on TQ; lifting fields from Q to TQ We recall first that a differentiable vector field on Q is represented in a coordinate system $q = (q^1, \ldots, q^n)$ by n first-order ordinary differential equations

$$\frac{dq^i}{d\epsilon} = f^i(q^1, \ldots, q^n). \tag{3.1}$$

A vector field generates a one-parameter family of active transformations: viz. passage along the vector field's integral curves, by a varying parameter-difference ϵ. The vector field is called the *infinitesimal generator* of the family. It is common to write the parameter as τ, but in this Section we use ϵ to avoid confusion with t, which often represents the time.

Similarly, a vector field defined on TQ corresponds to a system of $2n$ ordinary differential equations, and generates an active transformation of TQ. But I will consider only vector fields on TQ that mesh with the structure of TQ as a tangent bundle, in the sense that they are induced by vector fields on Q, in the following natural way.

This induction has two ingredient ideas.

First, any curve in Q (representing a possible state of motion) defines a corresponding curve in TQ, because the functions $q^i(t)$ define the functions $\dot{q}^i(t)$. (Here t is the parameter of the curve.) More formally: given any curve in configuration space, $\phi : I \subset \mathbb{R} \to Q$, with coordinate expression in the q-system $t \in I \mapsto q(\phi(t)) \equiv q(t) = q^i(t)$, we define its *extension* to TQ to be the curve $\Phi : I \subset \mathbb{R} \to TQ$ given in the corresponding coordinates by $q^i(t), \dot{q}^i(t)$.

Second, any vector field X on Q generates displacements in any possible state of motion, represented by a curve in Q with coordinate expression $q^i = q^i(t)$. Namely: for a given value of the parameter ϵ, the displaced state of motion is represented by the curve in Q

$$q^i(t) + \epsilon X^i(q^i(t)). \tag{3.2}$$

Putting these ingredients together: we first displace a curve within Q, and then extend the result to TQ. Namely, the extension to TQ of the (curve representing)

the displaced state of motion is given by the $2n$ functions, in two groups each of n functions, for the $(q, \dot{q})$ coordinate system

$$q^i(t) + \epsilon X^i(q^i(t)) \quad \text{and} \quad \dot{q}^i(t) + \epsilon Y^i(q^i(t), \dot{q}^i); \tag{3.3}$$

where Y is defined to be the vector field on TQ that is the derivative along the original state of motion of X. That is:

$$Y^i(q, \dot{q}) := \frac{dX^i}{dt} = \Sigma_j \frac{\partial X^i}{\partial q^j} \dot{q}^j. \tag{3.4}$$

Thus displacements by a vector field within Q are lifted to TQ. The vector field X on Q lifts to TQ as $(X, \frac{dX}{dt})$; i.e. it lifts to the vector field that sends a point $(q^i, \dot{q}^i) \in TQ$ to $\left(q^i + \epsilon X^i, \dot{q}^i + \epsilon \frac{dX^i}{dt}\right)$.[9]

3.2.2 The definition of variational symmetry To define variational symmetry, I begin with the integral notion and then give the differential notion. The idea is that the Lagrangian L, a scalar $L : TQ \to \mathbb{R}$, should be invariant under all the elements of a one-parameter family of active transformations $\theta_\epsilon : \epsilon \in I \subset \mathbb{R}$: at least in a neighbourhood of the identity map corresponding to $\epsilon = 0$, $\theta_0 \equiv id_U$. (Here U is some open subset of TQ, maybe not all of it.)

That is, we define the family $\theta_\epsilon : \epsilon \in I \subset \mathbb{R}$ to be a *variational symmetry* of L if L is invariant under the transformations: $L = L \circ \theta_\epsilon$, at least around $\epsilon = 0$. (We could use the correspondence between active and passive transformations to recast this definition, and what follows, in terms of a passive notion of symmetry as sameness of L's functional form in different coordinate systems. I leave this as an exercise! Or cf. Butterfield (2004a: Section 4.7.2).)

For the differential notion of variational symmetry, we of course use the idea of a vector field. But we also impose Section 3.2.1's restriction to vector fields on TQ that are induced by vector fields on Q. So we define a vector field X on Q that generates a family of active transformations θ_ϵ on TQ to be a variational symmetry of L if the first derivative of L with respect to ϵ is zero, at least around $\epsilon = 0$. More precisely: writing

$$L \circ \theta_\epsilon = L(q^i + \epsilon X^i, \dot{q}^i + \epsilon Y^i) \quad \text{with} \quad Y^i = \Sigma_j \frac{\partial X^i}{\partial q^j} \dot{q}^j, \tag{3.5}$$

we say X is a *variational symmetry* iff the first derivative of L with respect to ϵ is zero (at least around $\epsilon = 0$). That is: X is a variational symmetry iff

$$\Sigma_i \, X^i \frac{\partial L}{\partial q^i} \; + \; \Sigma_i \, Y^i \frac{\partial L}{\partial \dot{q}^i} \; = 0 \quad \text{with} \quad Y^i = \Sigma_j \frac{\partial X^i}{\partial q^j} \dot{q}^j. \tag{3.6}$$

3.2.3 A contrast with dynamical symmetries The general notion of a dynamical symmetry, i.e. a symmetry of some equations of motion (whether Euler-Lagrange or not), is not needed for Section 3.4's presentation of Noether's theorem. But the notion is so important that I must mention it, though only to contrast it with variational symmetries.

The general definition is roughly as follows. Given any system of differential equations, $\mathcal{E}$ say, a *dynamical symmetry* of the system is an active transformation ζ on the system $\mathcal{E}$'s space of both independent variables, x_j say, and dependent variables y^i say, such that any solution of $\mathcal{E}$, $y^i = f^i(x_j)$ say, is carried to another solution. For a precise definition, cf. Olver (2000: Def. 2.23, p. 93), and his ensuing discussion of the induced action (called 'prolongation') of the transformation ζ on the spaces of (in general, partial) derivatives of the y's with respect to the x's (i.e. jet spaces).

As I said in Section 3.1, groups of symmetries in this sense play a central role in the modern theory of differential equations: not just in finding new solutions, once given a solution, but also in integrating the equations. For some main theorems stating criteria (in terms of prolongations) for groups of symmetries, cf. Olver (2000: Theorem 2.27, p. 100, Theorem 2.36, p. 110, Theorem 2.71, p. 161).

But for present purposes, it is enough to state the rough idea of a one-parameter group of dynamical symmetries (without details about prolongations!) for Lagrange's equations in the familiar form, eq. 2.1.

In this simple case, there is just one independent variable $x := t$, so that:

(a) we are considering ordinary, not partial, differential equations, with n dependent variables $y^i := q^i(t)$.

(b) prolongations correspond to lifts of maps on Q to maps on TQ; cf. Section 3.2.1.

Furthermore, in line with the discussion following Lagrange's equations eq. 2.1, the time-independence of the Lagrangian (time being a cyclic coordinate) means we can define dynamical symmetries ζ in terms of active transformations on the tangent bundle, $\theta : TQ \to TQ$, that are lifted from active transformations on Q. In effect, we define such a map ζ by just adjoining to any such $\theta : TQ \to TQ$ the identity map on the time variable $id : t \in \mathbb{R} \mapsto t$. (More formally: $\zeta : (q, \dot{q}, t) \in TQ \times \mathbb{R} \mapsto (\theta(q, \dot{q}), t) \in TQ \times \mathbb{R}$.)

Then we define in the usual way what it is for a one-parameter family of such maps $\zeta_s : s \in I \subset \mathbb{R}$ to be a (local) one-parameter group of dynamical symmetries (for Lagrange's equations eq. 2.1): namely, if any solution curve $q(t)$ (equivalently: its extension $q(t), \dot{q}(t)$ to TQ) of the Lagrange equations is carried by each ζ_s to another solution curve, with the ζ_s for different s composing in the obvious way, for s close enough to $0 \in I$.

And finally: we also define (in a manner corresponding to the discussion at the end of Section 3.2.2) a differential, as against integral, notion of dynamical symmetry. Namely, we say a vector field X on Q is a dynamical symmetry if its lift to TQ (more precisely: its lift, with the identity map on the time variable adjoined) is the infinitesimal generator of such a one-parameter family ζ_s.

For us, the important point is that this notion of a dynamical symmetry is *different* from Section 3.2.2's notion of a variational symmetry.[10] As I announced in Section 3.1,

a variational symmetry is (with one *caveat*) a dynamical symmetry—but the converse is false. Fortunately, the same simple example will serve both to show the subtlety about the first implication, and as a counterexample to the converse implication. This example is the two-dimensional harmonic oscillator.[11]

The usual Lagrangian is, with cartesian coordinates written as qs, and the contravariant indices written for clarity as subscripts:

$$L_1 = \frac{1}{2}\left[\dot{q}_1^2 + \dot{q}_2^2 \; - \; \omega^2(q_1^2 + q_2^2)\right]; \tag{3.7}$$

giving as Lagrange equations:

$$\ddot{q}_i + \omega^2 q_i = 0, \;\; i = 1, 2. \tag{3.8}$$

But these Lagrange equations, i.e. the same dynamics, are also given by

$$L_2 = \dot{q}_1\dot{q}_2 - \omega^2 q_1 q_2. \tag{3.9}$$

The rotations in the plane are of course a variational symmetry of L_1, and a dynamical symmetry of eq. 3.8. But they are *not* a variational symmetry of L_2. So a dynamical symmetry need not be a variational one. Besides, these equations contain another example to the same effect. Namely, the "squeeze" transformations

$$q_1' := e^{\eta} q_1, \; q_2' := e^{-\eta} q_2 \tag{3.10}$$

are a dynamical symmetry of eq. 3.8, but not a variational symmetry of L_1. So again: a dynamical symmetry need not be a variational one.[12]

I turn to the first implication: that every variational symmetry is a dynamical symmetry. This is true: general and abstract proofs (applying also to continuous systems i.e. field theories) can be found in Olver (2000: theorem 4.14, p. 255; theorem 4.34, p. 278; theorem 5.53, p. 332).

But beware of a condition of the theorem. (This is the *caveat* mentioned at the end of Section 3.1.) The theorem requires that all the variables q (for continuous systems: all the fields ϕ) be subject to Hamilton's Principle. The need for this condition is shown by rotations in the plane, which are a variational symmetry of the familiar Lagrangian L_1 above. But it is easy to show that such a rotation is a dynamical symmetry of one of the Lagrange equations, say the equation for the variable q_1

$$\ddot{q}_1 + \omega^2 q_1 = 0, \tag{3.11}$$

only if the corresponding Lagrange equation holds for q_2.

3.3 The conjugate momentum of a vector field

Now we define *the momentum conjugate to a vector field* X to be the scalar function on TQ:

$$p_X : TQ \to \mathbb{R} \;\; ; \;\; p_X = \Sigma_i \, X^i \frac{\partial L}{\partial \dot{q}^i} \tag{3.12}$$

(For a time-dependent Lagrangian, p_X would be a scalar function on $TQ \times \mathbb{R}$, with $\mathbb{R}$ representing time.)

We shall see in the next Subsection's examples that this definition generalizes in an appropriate way Section 2.1's definition of the momentum conjugate to a coordinate q.

But first note that it is an *improvement* in the sense that, while the momentum conjugate to a coordinate q depends on the choice made for the other coordinates, the momentum p_X conjugate to a vector field X is independent of the coordinates chosen. Though this point is not needed in order to prove Noether's theorem, here is the proof.

We first apply the chain-rule to $L = L(q'(q), \dot{q}'(q, \dot{q}))$ and eq. 2.11 ("cancellation of the dots"), to get

$$\frac{\partial L}{\partial \dot{q}^i} = \Sigma_j \frac{\partial L}{\partial \dot{q}'^j} \frac{\partial \dot{q}'^j}{\partial \dot{q}^i} = \Sigma_j \frac{\partial L}{\partial \dot{q}'^j} \frac{\partial q'^j}{\partial q^i}. \tag{3.13}$$

Then using the transformation law for components of a vector field

$$X'^i = \Sigma_j \frac{\partial q'^i}{\partial q^j} X^j. \tag{3.14}$$

and relabelling i and j, we deduce:

$$\begin{aligned} p'_X &= \Sigma_i X'^i \frac{\partial L}{\partial \dot{q}'^i} \\ &= \Sigma_{ij} X^j \frac{\partial q'^i}{\partial q^j} \frac{\partial L}{\partial \dot{q}'^i} = \Sigma_{ij} X^i \frac{\partial q'^j}{\partial q^i} \frac{\partial L}{\partial \dot{q}'^j} = \Sigma_i X^i \frac{\partial L}{\partial \dot{q}^i} \equiv p_X. \end{aligned} \tag{3.15}$$

Finally, I remark incidentally that in the geometric formulation of Lagrangian mechanics (Section 2.2) , the coordinate-independence of p_X becomes, unsurprisingly, a triviality. Namely: p_X is obviously the contraction of X as lifted to TQ with the canonical1-form on TQ that we defined in eq. 2.13:

$$\theta_L := \frac{\partial L}{\partial \dot{q}^i} dq^i. \tag{3.16}$$

We will return to this at the end of Section 3.4.1.

3.4 Noether's theorem; and examples

Given just the definition of conjugate momentum, eq. 3.12, the proof of Noether's theorem isimmediate. (The interpretation and properties of this momentum, discussed in the last Subsection, are not needed.) The theorem says:

Noether's theorem for Lagrangian mechanics If X is a (variational) symmetry of a system with Lagrangian $L(q, \dot{q}, t)$, then X's conjugate momentum is a constant of the motion.

Proof: We just calculate the derivative of the momentum eq. 3.12 along the solution curves in TQ, and apply Lagrange's equations and the definitions of Y^i, and of symmetry eq. 3.6:

$$\begin{aligned} \frac{dp}{dt} &= \Sigma_i \, \frac{dX^i}{dt} \frac{\partial L}{\partial \dot{q}^i} \; + \; \Sigma_i \, X^i \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{q}^i} \right) \\ &= \; \Sigma_i \, Y^i \frac{\partial L}{\partial \dot{q}^i} \; + \; \Sigma_i \, X^i \frac{\partial L}{\partial q^i} \; = 0. \quad \text{QED.} \end{aligned} \tag{3.17}$$

Examples: This proof, though neat, is a bit abstract! So here are two examples, both of which return us to examples we have already seen.

(1) The first example is a shift in a cyclic coordinate q^n: i.e. the case with which our discussion of Noether's theorem began at the end of Section 2.1. So suppose q^n is cyclic, and define a vector field X by

$$X^1 = 0, \ldots, X^{n-1} = 0, \; X^n = 1. \tag{3.18}$$

So the displacements generated by X are translations by an amount ϵ in the q^n-direction. Then $Y^i := \frac{dX^i}{dt}$ vanishes, and the definition of (variational) symmetry eq. 3.6 reduces to

$$\frac{\partial L}{\partial q^n} = 0. \tag{3.19}$$

So since q^n is assumed to be cyclic, X is a symmetry. And the momentum conjugate to X, which Noether's theorem tells us is a constant of the motion, is the familiar one:

$$p_X := \Sigma_i \, X^i \frac{\partial L}{\partial \dot{q}^i} = \frac{\partial L}{\partial \dot{q}^n}. \tag{3.20}$$

As mentioned in Section 3.1, this example is *universal*, in that every symmetry X arises, around any point where X is non-zero, from a cyclic coordinate in some local system of coordinates. This follows from the basic theorem about the local existence and uniqueness of solutions of ordinary differential equations. We can state the theorem as follows; (cf. e.g. Arnold (1973: 48–49, 77–78, 249–250), Olver (2000: Prop 1.29)).

Consider a system of n first-order ordinary differential equations on an open subset U of an n-dimensional manifold

$$\dot{q^i} = X^i(q) \equiv X^i(q^1, \ldots, q^n), \quad q \in U; \tag{3.21}$$

equivalently, a vector field X on U. Let q_0 be a non-singular point of the vector field, i.e. $X(q_0) \neq 0$. Then in a sufficiently small neighbourhood V of q_0, there is a coordinate system (formally, a diffeomorphism $f : V \to W \subset \mathbb{R}^n$) such that, writing

$y_i : \mathbb{R}^n \to \mathbb{R}$ for the standard coordinates on W and $\mathbf{e}_i$ for the ith standard basis vector of $\mathbb{R}^n$, eq. 3.21 goes into the very simple form

$$\dot{\mathbf{y}} = \mathbf{e}_n; \text{ i.e. } \dot{y}_n = 1, \ \dot{y}_1 = \dot{y}_2 = \cdots = \dot{y}_{n-1} = 0 \text{ in } W. \tag{3.22}$$

(In terms of the tangent map (also known as: push-forward) f_* on tangent vectors that is induced by f: $f_*(X) = \mathbf{e}_n$ in W.) On account of eq. 3.22's simple form, Arnold suggests the theorem might well be called the 'rectification theorem'.

We should note two points about the theorem:

(i) The rectifying coordinate system f may of course be very hard to find. So the theorem by no means makes all problems "trivially soluble"; cf. again footnote 4.

(ii) The theorem has an immediate corollary about *local* constants of the motion. Namely: n first-order ordinary differential equations have, locally, $n-1$ functionally independent constants of the motion (also known as: first integrals). They are given, in the above notation, by $y_1, \ldots, y_{n-1}$.

We now apply the rectification theorem, so as to reverse the reasoning in the above example of q^n cyclic. That is: assuming X is a symmetry, let us rectify it—i.e. let us pass to a coordinate system (q) such that eq. 3.18 holds. Then, as above, $Y^i := \frac{dX^i}{dt}$ vanishes; and X's being a (variational) symmetry, eq. 3.6, reduces to q^n being cyclic; and the momentum conjugate to X, p_X reduces to the familiar conjugate momentum $p_n = \frac{\partial L}{\partial \dot{q}^n}$. Thus every symmetry X arises locally from a cyclic coordinate q^n and the corresponding conserved momentum is p_n. (But note that this may hold only "very locally": the domain V of the coordinate system f in which X generates displacements in the direction of the cyclic coordinate q^n can be smaller than the set U on which X is a symmetry.)

In Section 5.3, the fact that every symmetry arises locally from a cyclic coordinate will be important for understanding the Hamiltonian version of Noether's theorem.

(2) Let us now look at our previous example, the angular momentum of a free particle (eq. 2.6), in the *cartesian* coordinate system, i.e. a coordinate system *without* cyclic coordinates. So let $q_1 := x, q_2 := y, q_3 := z$. (In this example, subscripts will again be a bit clearer.) Then a small rotation about the x-axis

$$\delta x = 0, \ \ \delta y = -\epsilon z, \ \ \delta z = \epsilon y \tag{3.23}$$

corresponds to a vector field X with components

$$X_1 = 0, \ \ X_2 = -q_3, \ \ X_3 = q_2 \tag{3.24}$$

so that the Y_i are

$$Y_1 = 0, \ \ Y_2 = -\dot{q}_3, \ \ X_3 = \dot{q}_2. \tag{3.25}$$

For the Lagrangian

$$L = \frac{1}{2} m \left(\dot{q}_1^2 + \dot{q}_2^2 + \dot{q}_3^2 \right) \tag{3.26}$$

X is a (variational) symmetry since the definition of symmetry eq. 3.6 now reduces to

$$\Sigma_i \, X_i \frac{\partial L}{\partial q_i} \; + \; \Sigma_i \, Y_i \frac{\partial L}{\partial \dot{q}_i} = -\dot{q}_3 \frac{\partial L}{\partial \dot{q}_2} + \dot{q}_2 \frac{\partial L}{\partial \dot{q}_3} = 0. \tag{3.27}$$

So Noether's theorem then tells us that X's conjugate momentum is

$$p_X := \Sigma_i \, X_i \frac{\partial L}{\partial \dot{q}_i} = X_2 \frac{\partial L}{\partial \dot{q}_2} + X_3 \frac{\partial L}{\partial \dot{q}_3} = -mz\dot{y} + my\dot{z} \tag{3.28}$$

which is indeed the x-component of angular momentum.

3.4.1 A geometrical formulation We can give a geometric formulation of Noether's theorem by using the vanishing of the Lie derivative to express constancy along the integral curves of a vector field. There are two vector fields on TQ to consider: the dynamical vector field D (cf. eq. 2.8), and the lift to TQ of the vector field X that is the variational symmetry.

I will now write $\bar{X}$ for this lift. So given the vector field X on Q

$$X = X^i(q) \frac{\partial}{\partial q^i}, \tag{3.29}$$

the lift $\bar{X}$ of X to TQ is, by eq. 3.4,

$$\bar{X} = X^i(q) \frac{\partial}{\partial q^i} + \frac{\partial X^i(q)}{\partial q^j} \dot{q}^j \frac{\partial}{\partial \dot{q}^i}, \tag{3.30}$$

where the q argument of X^i emphasises that the X^i do not depend on $\dot{q}$.

That X is a variational symmetry means that in TQ, the Lie derivative of L along the lift $\bar{X}$ vanishes: $\mathcal{L}_{\bar{X}} L = 0$. On the other hand, we know from eq. 3.16 that the momentum p_X conjugate to X is the contraction $<;>$ of $\bar{X}$ with the canonical 1-form $\theta_L := \frac{\partial L}{\partial \dot{q}^i} dq^i$ on TQ:

$$p_X \; := \; X^i \frac{\partial L}{\partial \dot{q}^i} \; \equiv < \bar{X}; \theta_L > . \tag{3.31}$$

So Noether's theorem says:

$$\textit{If } \mathcal{L}_{\bar{X}} L = 0, \text{ then } \mathcal{L}_D < \bar{X}; \theta_L >= 0.$$

Note finally that eq. 3.31 shows that the theorem has no converse. That is: given that a dynamical variable $p : TQ \to \mathbb{R}$ is a constant of the motion, $\mathcal{L}_D p = 0$, there is no single vector field $\bar{X}$ on TQ such that $p =< \bar{X}; \theta_L >$. For given such a $\bar{X}$, one could get another by adding any field $\bar{Y}$ for which $< \bar{Y}; \theta_L >= 0$. However, we will see in Section 5.2 that in Hamiltonian mechanics a constant of the motion *does* determine a corresponding vector field on the state space.

4 HAMILTONIAN MECHANICS INTRODUCED

4.1 Preamble

From now on this paper adopts the Hamiltonian framework. As we shall see, its description of symmetry and conserved quantities is in various ways more straightforward and powerful than that of the Lagrangian framework.

The main idea is to replace the $\dot{q}$s by the canonical momenta, the ps. More generally, the state-space is no longer the tangent bundle TQ but a phase space Γ, which we take to be the cotangent bundle T^*Q. (Here, the phrase 'we take to be' just signals the fact that eventually, in Section 6.8, we will glimpse a more general kind of Hamiltonian state-space, viz. Poisson manifolds.)

Admittedly, the theory on TQ given by Lagrange's equations eq. 2.1 is equivalent to the Hamiltonian theory on T^*Q given by eq. 4.5 below, once we assume the Hessian condition eq. 2.3.

But of course, theories can be formally equivalent, but different as regards their power for solving problems, their heuristic value and even their interpretation. In our case, two advantages of Hamiltonian mechanics over Lagrangian mechanics are commonly emphasised. (i) The first concerns its greater power or flexibility for describing a given system, that Lagrangian methods can also describe (and so its greater power for solving problems about such a system). (ii) The second concerns the broader idea of describing other systems. In more detail:

(i) Hamiltonian mechanics replaces the group of *point transformations*, $q \to q'$ on Q, together with their lifts to TQ, by a "corresponding larger" group of transformations on Γ, the group of *canonical transformations* (also known as, for the standard case where $\Gamma = T^*Q$: the *symplectic group*).

This group "corresponds" to the point transformations (and their lifts) in that while for any Lagrangian L, Lagrange's equations eq. 2.1 are covariant under all the point transformations, Hamilton's equations eq. 4.5 below are (for any Hamiltonian H) covariant under all canonical transformations. And it is a "larger" group because:

(a) any point transformation together with its lift to TQ is a canonical transformation: (more precisely: it naturally defines a canonical transformation on T^*Q);

(b) not every canonical transformation is thus induced by a point transformation; for a canonical transformation can "mix" the qs and ps in a way that point transformations and their lifts cannot.

There is a rich and multi-faceted theory of canonical transformations, to which there are three main approaches—generating functions, integral invariants and symplectic geometry. I will adopt the symplectic approach, but not need many details about it. In particular, we will need only a few details about how the "larger" group of canonical transformations makes for a more powerful version of Noether's theorem.

(ii) The Hamiltonian framework connects analytical mechanics with other fields of physics, especially statistical mechanics and optics. The first connection goes via canonical transformations, especially using the integral invariants approach. The second connection goes via Hamilton-Jacobi theory; (for a philosopher's

exposition, with an eye on quantum theory, cf. Butterfield (2004b: especially Sections 7–9)).[13]

With its theme of symmetry and conservation, this paper will illustrate (i), greater power in describing a given system, rather than (ii), describing other systems. As to (i), we will see two main ways in which the Hamiltonian framework is more powerful than the Lagrangian one. First, cyclic coordinates will "do more work for us" (Section 4.2). Second, the Hamiltonian version of Noether's theorem is both: more powerful, thanks to the use of the "larger" group of canonical transformations; and more easily proven, thanks to the use of Poisson brackets (Section 5).

So from now on, the broad plan is as follows. After Section 4.2's deduction of Hamilton's equations, Section 4.3 introduces symplectic structure, starting from the "naive" form of the symplectic matrix. Section 5 presents Poisson brackets, and the Hamiltonian version of Noether's theorem. Finally, Section 6 gives a geometric perspective, corresponding to Section 2.2's geometric perspective on the Lagrangian framework.

4.2 Hamilton's equations

4.2.1 The equations introduced Recall the vision in (5) of Section 2.2.2: that we seek $2n$ new variables, ξ^α say, $\alpha = 1, \ldots, 2n$ in which Lagrange's equations take the simple form

$$\frac{d\xi^\alpha}{dt} = f_\alpha(\xi^1, \ldots \xi^{2n}). \tag{4.1}$$

We can find the desired variables ξ^α by using the canonical momenta

$$p_i := \frac{\partial L}{\partial \dot{q}^i} =: L_{\dot{q}^i}, \tag{4.2}$$

to write the $2n$ Lagrange equations as

$$\frac{dp_i}{dt} = \frac{\partial L}{dq^i}; \ \frac{dq^i}{dt} = \dot{q}^i. \tag{4.3}$$

These are of the desired simple form, except that the right hand sides need to be written as functions of (q, p, t) rather than $(q, \dot{q}, t)$. (Here and in the next two paragraphs, we temporarily allow time-dependence, since the deduction is unaffected: the time variable is "carried along unaffected". In the terms of Section 2.1, this means allowing non-scleronomous constraints and a time-dependent work-function U.)

For the second group of n equations, this is in principle straightforward, given our assumption of a non-zero Hessian, eq. 2.3. This implies that we can invert eq. 4.2 so as to get the n $\dot{q}^i$ as functions of (q, p, t). We can then apply this to the first group of equations; i.e. we substitute $\dot{q}^i(q, p, t)$ wherever $\dot{q}^i$ appears in any right hand side $\frac{\partial L}{dq^i}$.

But we need to be careful: the partial derivative of $L(q,\dot{q},t)$ with respect to q^i is not the same as the partial derivative of $\hat{L}(q,p,t) := L(q,\dot{q}(q,p,t),t)$ with respect to q^i, since the first holds fixed the $\dot{q}$s, while the second holds fixed the ps. A comparison of these partial derivatives leads, with algebra, to the result that if we define the *Hamiltonian function* by

$$H(q,p,t) := p_i\dot{q}^i(q,p,t) - \hat{L}(q,p,t) \tag{4.4}$$

then the $2n$ equations eq. 4.3 go over to *Hamilton's equations*

$$\frac{dp_i}{dt} = -\frac{\partial H}{\partial q^i}; \quad \frac{dq^i}{dt} = \frac{\partial H}{\partial p_i}. \tag{4.5}$$

So we have cast our $2n$ equations in the simple form, $\frac{d\xi^\alpha}{dt} = f_\alpha(\xi^1,\ldots\xi^{2n})$, requested in (5) of Section 2.2. More explicitly: defining

$$\xi^\alpha = q^\alpha, \;\; \alpha = 1,\ldots,n; \quad \xi^\alpha = p_{\alpha-n}, \;\; \alpha = n+1,\ldots,2n \tag{4.6}$$

Hamilton's equations become

$$\dot{\xi}^\alpha = \frac{\partial H}{\partial \xi^{\alpha+n}}, \;\; \alpha = 1,\ldots,n; \quad \dot{\xi}^\alpha = -\frac{\partial H}{\partial \xi^{\alpha-n}}, \;\; \alpha = n+1,\ldots,2n. \tag{4.7}$$

To sum up: a single function H determines, through its partial derivatives, the evolution of all the qs and ps—and so, the evolution of the state of the system.

4.2.2 Cyclic coordinates in the Hamiltonian framework Just from the form of Hamilton's equations, we can immediately see a result that is significant for our theme of how symmetries and conserved quantities reduce the number of variables involved in a problem. In short, we can see that with Hamilton's equations in hand, cyclic coordinates will "do more work for us" than they do in the Lagrangian framework.

More specifically, recall the basic Lagrangian result from the end of Section 2.1, that the generalized momentum $p_n := \frac{\partial L}{\partial \dot{q}^n}$ is conserved if, indeed iff, its conjugate coordinate q^n is cyclic, $\frac{\partial L}{\partial q^n} = 0$. And recall from Section 3.4 that this result underpinned Noether's theorem in the precise sense of being "universal" for it. Corresponding results hold in the Hamiltonian framework—but are in certain ways more powerful.

Thus we first observe that the transformation "from the $\dot{q}$s to the ps", i.e. the transition between Lagrangian and Hamiltonian frameworks, does not involve the dependence on the qs. More precisely: partially differentiating eq. 4.4 with respect to q^n, we obtain

$$\frac{\partial H}{\partial q^n} \equiv \frac{\partial H}{\partial q^n}\,|_{p;q^i,i\neq n} \;\; = \;\; -\frac{\partial L}{\partial q^n} \equiv -\frac{\partial L}{\partial q^n}\,|_{\dot{q};q^i,i\neq n}\,. \tag{4.8}$$

(The other two terms are plus and minus $p_i \frac{\partial \dot{q}^i}{\partial q^n}$, and so cancel.) So a coordinate q^n that is cyclic in the Lagrangian sense is also cyclic in the obvious Hamiltonian sense, viz. that $\frac{\partial H}{\partial q^n} = 0$. But by Hamilton's equations, this is equivalent to $\dot{p}_n = 0$. So we have the result corresponding to the Lagrangian one: p_n is conserved iff q_n is cyclic (in the Hamiltonian sense).

We will see in Section 5.3 that this result underpins the Hamiltonian version of Noether's theorem; just as the corresponding Lagrangian result underpinned the Lagrangian version of Noether's theorem (cf. discussion after eq. 3.20).

But we can already see that this result gives the Hamiltonian formalism an advantage over the Lagrangian. In the latter, the generalized velocity corresponding to a cyclic coordinate, q_n will in general still occur in the Lagrangian. The Lagrangian will be $L(q_1, \ldots, q_{n-1}, \dot{q}_1, \ldots, \dot{q}_n, t)$, so that we still face a problem in n variables.

But in the Hamiltonian formalism, p_n will be a constant of the motion, α say, so that the Hamiltonian will be $H(q_1, \ldots, q_{n-1}, p_1, \ldots, p_{n-1}, \alpha, t)$. So we now face a problem in $n-1$ variables, α being simply determined by the initial conditions. That is: after solving the problem in $n-1$ variables, q_n is determined just by quadrature: i.e. just by integrating (perhaps numerically) the equation

$$\dot{q}_n = \frac{\partial H}{\partial \alpha}, \tag{4.9}$$

where, thanks to having solved the problem in $n-1$ variables, the right-hand side is now an explicit function of t.

This result is very simple. But it is an important illustration of the power of the Hamiltonian framework. Indeed, Arnold remarks (1989: 68) that 'almost all the solved problems in mechanics have been solved by means of' it!

No doubt his point is, at least in part, that this result underpins the Hamiltonian version of Noether's theorem. But I should add that the result also motivates the study of various notions related to the idea of cyclic coordinates, such as constants of the motion being in involution (i.e. having zero Poisson bracket with each other), and a system being completely integrable (in the sense of Liouville). These notions have played a large part in the way that Hamiltonian mechanics has developed, especially in its theory of canonical transformations. And they play a large part in the way Hamiltonian mechanics has solved countless problems. But as announced in Section 4.1, this paper will not go into these aspects of Hamiltonian mechanics, since they are not needed for our theme of symmetry and conservation; (for a philosophical discussion of these aspects, cf. Butterfield 2005).

4.2.3 The Legendre transformation and variational principles To end this Subsection, I note two aspects of this transition from Lagrange's equations to Hamilton's. For, although I shall not need details about them, they each lead to a rich theory:

(i) The transformation "from the $\dot{q}$s to the ps" is the *Legendre transformation*. It has a striking geometric interpretation. In the simplest case, it concerns the fact that one can describe a smooth convex real function $y = f(x), f''(x) > 0$, not by the pairs

of its arguments and values (x, y), but by the pairs of its gradients at points (x, y) and the intercepts of its tangent lines with the y-axis. Given the non-zero Hessian (eq. 2.3), one readily proves various results: e.g. that the geometric interpretation extends to higher dimensions, and that the transformation is self-inverse, i.e. its square is the identity. For details, cf. e.g.: Arnold (1989: Chapters 3.14, 9.45.C), Courant and Hilbert (1953: Chapter IV.9.3; 1962, Chapter I.6), José and Saletan (1998: 212–217), Lanczos (1986: Chapter VI.1-4). The Legendre transformation is also described using modern geometry's idea of a *fibre derivative*; as we will see briefly in Section 6.7.

(ii) The transition to Hamilton's equations has achieved more than we initially sought with our eq. 4.1. Namely: all the f_α, all the right hand sides in Hamilton's equations, are up to a sign, partial derivatives of a single function H. In the Hamiltonian framework, it is precisely this feature that underpins the possibility of expressing the equations of motion by variational principles; (of course, the Lagrangian framework has a corresponding feature). But as I mentioned, this paper does not discuss variational principles; for details cf. e.g. Lanczos (1986: Chapter VI.4) and Butterfield (2004: especially Section 5.2).

To sum up this introduction to Hamilton's equations:— Even once we set aside (i) and (ii), these equations mark the beginning of a rich and multi-faceted theory. At the centre lies the $2n$-dimensional phase space Γ coordinatized by the qs and ps: or more precisely, as we shall see later, the cotangent bundle T^*Q. The structure of Hamiltonian mechanics is encoded in the structure of Γ, and thereby in the coordinate transformations on Γ that preserve this structure, especially the form of Hamilton's equations: the canonical transformations. As I mentioned in Section 4.1, these transformations can be studied from three main perspectives: generating functions, integral invariants and symplectic structure—but I shall only need the last.

4.3 Symplectic forms on vector spaces

I shall introduce symplectic structure by giving Hamilton's equations a yet more symmetric appearance. This will lead to some elementary ideas about area in $\mathbb{R}^m$ and symplectic forms on vector spaces: ideas which will later be "made local" by taking the relevant copy of $\mathbb{R}^m$ to be the tangent space at a point of a manifold. (As usually formulated, Hamiltonian mechanics is especially concerned with the case $m = 2n$.)

4.3.1 Time-evolution from the gradient of H Writing $\mathbf{1}$ and $\mathbf{0}$ for the $n \times n$ identity and zero matrices respectively, we define the $2n \times 2n$ *symplectic matrix* ω by

$$\omega := \begin{pmatrix} \mathbf{0} & \mathbf{1} \\ -\mathbf{1} & \mathbf{0} \end{pmatrix}. \tag{4.10}$$

ω is antisymmetric, and has the properties, writing $\tilde{}$ for the transpose of a matrix, that

$$\tilde{\omega} = -\omega = \omega^{-1} \text{ so that } \quad \omega^2 = -\mathbf{1}; \text{ also } \quad \det \omega = 1. \tag{4.11}$$

Using ω, Hamilton's equations eq. 4.7 get the more symmetric form, in matrix notation

$$\dot{\xi} = \omega \frac{\partial H}{\partial \xi}. \tag{4.12}$$

In terms of components, writing $\omega^{\alpha\beta}$ for the matrix elements of ω, and defining $\partial_\alpha := \partial/\partial\xi^\alpha$, eq. 4.7 become

$$\dot{\xi}^\alpha = \omega^{\alpha\beta}\partial_\beta H. \tag{4.13}$$

Eq. 4.12 and 4.13 show how ω forms, from the naive gradient (column vector) ∇H of H on the phase space Γ of qs and ps, the vector field on Γ that gives the system's evolution: the *Hamiltonian vector field*, often written X_H. At a point $z = (q,p) \in \Gamma$, eq. 4.12 can be written

$$X_H(z) = \omega \nabla H(z). \tag{4.14}$$

The vector field X_H is also written as D (for 'dynamics'), on analogy with the Lagrangian framework's vector field D of eq. 2.8 in Section 2.2.

In Section 6, we will see how this definition of a *vector* field from a gradient, i.e. a *covector* or 1-form field, arises from Γ's being a cotangent bundle. More precisely, we will see that any cotangent bundle has an intrinsic symplectic structure that provides, at each point of the base-manifold, a natural i.e. basis-independent isomorphism between the tangent space and the cotangent space. For the moment, we:

(i) note a geometric interpretation of ω in terms of area (Section 4.3.2); and then
(ii) generalize the above discussion of ω into the definition of a symplectic form for a fixed vector space (Section 4.3.3).

4.3.2 Interpretation in terms of areas Let us begin with the simplest possible case: $\mathbb{R}^2 \ni (q,p)$, representing the phase space of a particle constrained to one spatial dimension. Here, the 2×2 matrix

$$\omega := \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \tag{4.15}$$

defines the antisymmetric bilinear form on $\mathbb{R}^2$:

$$A : ((q^1,p_1),(q^2,p_2)) \in \mathbb{R}^2 \times \mathbb{R}^2 \mapsto q^1p_2 - q^2p_1 \in \mathbb{R} \tag{4.16}$$

since

$$q^1p_2 - q^2p_1 = \begin{pmatrix} q^1 & p_1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} q^2 \\ p_2 \end{pmatrix} = \det \begin{pmatrix} q^1 & q^2 \\ p_1 & p_2 \end{pmatrix}. \tag{4.17}$$

It is easy to prove that $A((q^1,p_1),(q^2,p_2)) \equiv q^1p_2 - q^2p_1$ is the signed area of the parallelogram spanned by $(q^1,p_1),(q^2,p_2)$, where the sign is positive (negative) if the shortest rotation from (q^1,p_1) to (q^2,p_2) is anti-clockwise (clockwise).

Similarly in $\mathbb{R}^{2n}$: the matrix ω of eq. 4.10 defines an antisymmetric bilinear form on $\mathbb{R}^{2n}$ whose value on a pair $(q,p) \equiv (q^1,\ldots q^n;p_1,\ldots,p_n),(q',p') \equiv (q'^1,\ldots q'^n;p'_1,\ldots,p'_n)$ is the sum of the signed areas of the n parallelograms formed by the projections of the vectors $(q,p),(q',p')$ onto the n pairs of coordinate planes labelled $1,\ldots,n$. That is to say, the value is:

$$\sum_{i=1}^{n} q^ip'_i - q'^ip_i. \tag{4.18}$$

This induction of bilinear forms from antisymmetric matrices can be generalized: there is a one-to-one correspondence between forms and matrices. In more detail: there is a one-to-one correspondence between antisymmetric bilinear forms on $\mathbb{R}^2$ and antisymmetric 2×2 matrices. It is easy to check that any such form, ω say, is given, for any basis v,w of $\mathbb{R}^2$, by the matrix $\left(\begin{smallmatrix} 0 & \omega(v,w) \\ -\omega(v,w) & 0 \end{smallmatrix}\right)$. Similarly for any integer n: one easily shows that there is a one-to-one correspondence between antisymmetric bilinear forms on $\mathbb{R}^n$ and antisymmetric $n \times n$ matrices. (In Hamiltonian mechanics as usually formulated, we consider the case where n is even and the matrix is non-singular, as in eq. 4.10. But when one generalizes to Poisson manifolds (cf. Section 6.8) one allows n to be odd, and the matrix to be singular.)

This geometric interpretation of ω is important for two reasons.

(i) The first reason is that the idea of an antisymmetric bilinear form on a copy of $\mathbb{R}^{2n}$ is the main part of the definition of a symplectic form, which is the central notion in the usual geometric formulation of Hamiltonian mechanics. More details in Section 4.3.3, for a fixed copy of $\mathbb{R}^{2n}$; and in Section 6, where the form is defined on many copies of $\mathbb{R}^{2n}$, each copy being the tangent space at a point in the cotangent bundle T^*Q.

(ii) The second reason is that the idea of (signed) area underpins the theory of forms (1-forms, 2-forms etc.): i.e. antisymmetric multilinear functions on products of copies of $\mathbb{R}^n$. And when these copies of $\mathbb{R}^n$ are copies of the tangent space at (one and the same) point in a manifold, these forms lead to the whole theory of integration on manifolds. One needs this theory in order to make rigorous sense of any integration on a manifold beyond the most elementary (i.e. line-integrals); so it is crucial for almost any mathematical or physical theory using manifolds. In particular, it is crucial for Hamiltonian mechanics. So no wonder the *maestro* says that 'Hamiltonian mechanics cannot be understood without differential forms' (Arnold 1989, p. 163).

However, it turns out that this paper will not need many details about forms and the theory of integration. This is essentially because we focus only on solving mechanical problems, and simplifying them by appeals to symmetry. This means we will focus on line-integrals: viz. integrating with respect to time the equations of motion; or equivalently, integrating the dynamical vector field on the state space. We have already seen this vector field as X_H in eq. 4.14; and we will see it again, for example in terms

of Poisson brackets (eq. 5.14), and in geometric terms (Section 6). But throughout, the main idea will be as suggested by eq. 4.14: the vector field is determined by the symplectic matrix, "at" each point in the manifold Γ, acting on the gradient of the Hamiltonian function H.

So in short: focussing on line-integrals enables us to side-step most of the theory of forms.[14]

4.3.3 Bilinear forms and associated linear maps We now generalize from the symplectic matrix ω to a symplectic form; in five extended comments.

(1) *Preliminaries*:
Let V be a (real finite-dimensional) vector space, with basis $e_1, \ldots, e_i, \ldots e_n$. We write V^* for the dual space, and $e^1, \ldots, e^i, \ldots e^n$ for the dual basis: $e^i(e_j) := \delta^i_j$.

We recall that the isomorphism $e_i \mapsto e^i$ is basis-dependent: for a different basis, the corresponding isomorphism would be a different map. Only with the provision of appropriate extra structure would this isomorphism be basis-independent.

For physicists, the most familiar example of such a structure is the spacetime metric **g** in relativity theory. In terms of components, this basis-independence shows up in the way that **g** and its inverse lower and raise indices. As we will see in a moment, the underlying mathematical point is that because **g** is a bilinear form on a vector space V, i.e. $\mathbf{g} : V \times V \to \mathbb{R}$, and is non-degenerate, any $v \in V$ defines, independently of any choice of basis, an element of V^*: viz. the map $u \in V \mapsto \mathbf{g}(u, v)$. (In fact, V is the tangent space at a spacetime point; but this physical interpretation is irrelevant to the mathematical argument.) We will also see that Hamiltonian mechanics has a non-degenerate bilinear form, viz. a symplectic form, that similarly gives a basis-independent isomorphism between a vector space and its dual. (Roughly speaking, this vector space will be the $2n$-dimensional space of the qs and ps.)

On the other hand: for any vector space V, the isomorphism between V and V^{**} given by

$$e_i \mapsto [e_i] \in V^{**} : e^j \in V^* \mapsto e^j(e_i) = \delta^j_i \tag{4.19}$$

is basis-independent, and so we identify e_i with $[e_i]$, and V with V^{**}. We will write $< \, ; \, >$ (also written $< \, , \, >$) for the natural pairing (in either order) of V and V^*: e.g. $< e_i \, ; \, e^j > = < e^j \, ; \, e_i > = \delta^j_i$.

A linear map $A : V \to W$ induces (basis-independently) a *transpose* (aka: dual), written $\tilde{A}$ (or A^T or A^*), $\tilde{A} : W^* \to V^*$ by

$$\forall \alpha \in W^*, \forall v \in V : \quad \tilde{A}(\alpha)(v) \; \equiv < \tilde{A}(\alpha) \, ; \, v > := \; \alpha(A(v)) \; \equiv \; (\alpha \circ A)(v). \tag{4.20}$$

If $A : V \to W$ is a linear map between real finite-dimensional vector spaces, its matrix with respect to bases $e_1, \ldots, e_i, \ldots e_n$ and $f_1, \ldots, f_j, \ldots f_m$ of V and W is given by:

$$A(e_i) = A^j_i f_j; \quad \text{i.e. with } \; v = v^i e_i, \;\; (A(v))^j = A^j_i v^i. \tag{4.21}$$

So the upper index labels rows, and the lower index labels columns. Similarly, if $A : V \times W \to \mathbb{R}$ is a bilinear form, its matrix for these bases is defined as

$$A_{ij} := A(e_i, f_j) \tag{4.22}$$

so that on vectors $v = v^i e_i, w = w^j f_j$, we have: $A(v, w) = v^i A_{ij} w^j$.

(2) *Associated maps and forms*:
Given a bilinear form $A : V \times W \to \mathbb{R}$, we define the *associated linear map* $A^\flat : V \to W^*$ by

$$A^\flat(v)(w) \ := \ A(v, w). \tag{4.23}$$

Then $A^\flat(e_i) = A_{ij} f^j$: for both sides send any $w = w^j f_j$ to $A_{ij} w^j$. That is: the matrix of $A^\flat$ in the bases e_i, f^j of V and W^* is A_{ij}:

$$[A^\flat]_{ij} = A_{ij}. \tag{4.24}$$

On the other hand, we can proceed from linear maps to associated bilinear forms. Given a linear map $B : V \to W^*$, we define the *associated bilinear form* $B^\sharp$ on $V \times W^{**} \cong V \times W$ by

$$B^\sharp(v, w) \ = \ < B(v) \ ; \ w > . \tag{4.25}$$

If we put $A^\flat$ for B in eq. 4.25, its associated bilinear form, acting on vectors $v = v^i e_i, w = w^j f_j$, yields, by eq. 4.23:

$$(A^\flat)^\sharp(v, w) \ = \ < A^\flat(v); \ w > \ = \ A(v, w). \tag{4.26}$$

One similarly shows that if $B : V \to W^*$, then $\forall w \in W$:

$$\begin{aligned} (B^\sharp)^\flat(v)(w) &\equiv < (B^\sharp)^\flat(v); \ w > \ = \ B(v)(w) \\ &\equiv < B(v); w > \quad \text{so that} \quad (B^\sharp)^\flat = B. \end{aligned} \tag{4.27}$$

So the flat and sharp operations, $^\flat$ and $^\sharp$, are inverses.

(3) *Tensor products*:
It will sometimes be helpful to put the above ideas in terms of *tensor products*. If $v \in V, w \in W$, we can think of v and w as elements of V^{**}, W^{**} respectively. So we define their tensor product as a bilinear form on $V^* \times W^*$ by requiring for all $\alpha \in V^*, \beta \in W^*$:

$$(v \otimes w)(\alpha, \beta) \ := \ v(\alpha) w(\beta) \ \equiv < v \ ; \ \alpha >< w \ ; \ \beta > . \tag{4.28}$$

Similarly for other choices of vector spaces or their duals. Given $\alpha \in V^*, \beta \in W^*$, their tensor product is a bilinear form on $V \times W$:

$$(\alpha \otimes \beta)(v, w) \ := \ \alpha(v) \beta(w) \ \equiv < v \ ; \ \alpha >< w \ ; \ \beta > . \tag{4.29}$$

Similarly, we can think of $\alpha \in V^*, w \in W$ as elements of V^* and W^{**} respectively, and so define their tensor product as a bilinear form on $V \times W^*$:

$$(\alpha \otimes w)(v, \beta) \quad := \quad \alpha(v)w(\beta) \equiv < v\ ;\ \alpha >< w\ ;\ \beta > . \tag{4.30}$$

In this way we can express the linear map $A : V \to W$ in terms of tensor products. Since

$$A(e_i) = A^j_i f_j \quad \text{iff} \quad < A(e_i); f^j > \ = \ A^j_i \tag{4.31}$$

eq. 4.30 implies that

$$A \ = \ A^j_i \ e^i \otimes f_j. \tag{4.32}$$

Similarly, a bilinear form $A : V \times W \to \mathbb{R}$ with matrix $A_{ij} := A(e_i, f_j)$ (cf. eq. 4.22) is:

$$A \ = \ A_{ij} \ e^i \otimes f^j \tag{4.33}$$

The definitions of tensor product eq. 4.28, 4.29 and 4.30 generalize to higher-rank tensors (i.e. multilinear maps whose domains have more than two factors). But we will not need these generalizations.

(4) *Antisymmetric and non-degenerate forms*:
We now specialize to the forms and maps of central interest in Hamiltonian mechanics. We take $W = V$, dim(V)=n, and define a bilinear form $\omega : V \times V \to \mathbb{R}$ to be:
(i) *antisymmetric* iff: $\omega(v, v') = -\omega(v, v')$;
(ii) *non-degenerate* iff: if $\omega(v, v') = 0 \ \ \forall v' \in V$, then $v = 0$.
The form ω and its associated linear map $\omega^\flat : V \to V^*$ now have a square matrix ω_{ij} (cf. eq. 4.24). We define the *rank* of ω to be the rank of this matrix: equivalently, the dimension of the range $\omega^\flat(V)$.

We will also need the antisymmetrized version of eq. 4.29 that is definable when $W = V$. Namely, we define the *wedge-product* of $\alpha, \beta \in V^*$ to be the antisymmetric bilinear form on V, given by

$$\alpha \wedge \beta : (v, w) \in V \times V \mapsto (\alpha(v))(\beta(w)) - (\alpha(w))(\beta(v)) \in \mathbb{R} \ . \tag{4.34}$$

(The connection with Section 4.3.2, especially eq. 4.18, will become clear in a moment; and will be developed in Section 6.2.1.)

It is easy to show that for any bilinear form $\omega : V \times V \to \mathbb{R}$: ω is non-degenerate iff the matrix ω_{ij} is non-singular iff $\omega^\flat : V \to V^*$ is an isomorphism.

So a non-degenerate bilinear form establishes a basis-independent isomorphism between V and V^*; cf. the discussion of the spacetime metric **g** in (1) at the start of this Subsection.

Besides, this isomorphism $\omega^\flat$ has an inverse, suggesting another use of the sharp notation, viz. $\omega^\sharp$ is defined to be $(\omega^\flat)^{-1} : V^* \to V$. The isomorphism $\omega^\sharp : V^* \to V$

corresponds to ω's role, emphasised in Section 4.3.1, of defining a vector field X_H from dH. (But we will see in a moment that the space V implicitly considered in Section 4.3.1 had more structure than being just any finite-dimensional real vector space: viz. it was of the form $W \times W^*$.)

NB: This definition of $^\sharp$ is of course *not* equivalent to our previous definition, in eq. 4.25, since:

(i) on our previous definition, $^\sharp$ carried a linear map to a bilinear form, which reversed the passage by $^\flat$ from bilinear form to linear map, in the sense that for a bilinear form ω, we had $(\omega^\flat)^\sharp = \omega$; cf. eq. 4.26;

(ii) on the present definition, $^\sharp$ carries a bilinear form $\omega : V \times V \to \mathbb{R}$ to a linear map $\omega^\sharp : V^* \to V$, which inverts $^\flat$ in the sense (*different* from (i)) that

$$\omega^\sharp \circ \omega^\flat = id_V \text{ and } \omega^\flat \circ \omega^\sharp = id_{V^*}. \tag{4.35}$$

So beware: though not equivalent, both definitions are used! But it is a natural ambiguity, in so far as the definitions "mesh". For example, one easily shows that our second definition, i.e. eq. 4.35, is equivalent to a natural expression:

$$\forall \alpha, \beta \in V^* : \ < \omega^\sharp(\alpha), \beta > := \ \omega((\omega^\flat)^{-1}(\alpha), (\omega^\flat)^{-1}(\beta)). \tag{4.36}$$

It is also straightforward to show that for any bilinear form $\omega : V \times V \to \mathbb{R}$: if ω is antisymmetric of rank $r \leq n \equiv \dim(V)$, then r is even. That is: $r = 2s$ for some integer s, and there is a basis $e_1, \ldots, e_i, \ldots, e_n$ of V for which ω has a simple expansion as wedge-products

$$\omega \ = \ \sum_{i=1}^{s} e^i \wedge e^{i+s}; \tag{4.37}$$

equivalently, ω has the $n \times n$ matrix

$$\omega \ = \ \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{4.38}$$

where 1 is the $s \times s$ identity matrix, and similarly for the zero matrices of various sizes. This *normal form* of antisymmetric bilinear forms is an analogue of the Gram-Schmidt theorem that an inner product space has an orthonormal basis, and is proved by an analogous argument.

(5) *Symplectic forms*:
As usually formulated, Hamiltonian mechanics uses a non-degenerate antisymmetric bilinear form: i.e. $r = n$. So eq. 4.38 loses its bottom row and right column consisting of zero matrices, and reduces to the form of Section 4.3.1's naive symplectic matrix, eq. 4.10. Equivalently: eq. 4.37 reduces to eq. 4.18.

Accordingly, we define: a *symplectic form* on a (real finite-dimensional) vector space Z is a non-degenerate antisymmetric bilinear form ω on Z: $\omega : Z \times Z \to \mathbb{R}$. Z is then called a *symplectic vector space*. It follows that Z is of even dimension.

Besides, in Hamiltonian mechanics (as usually formulated) the vector space Z is a product $V \times V^*$ of a vector space and its dual. Indeed, this was already suggested by:

(i) the fact in (3) of Section 2.2.2, that the canonical momenta $p_i := \frac{\partial L}{\partial \dot{q}^i}$ transform as a 1-form, and

(ii) Section 4.3.1's discussion of the one-form field ∇H determining a vector field X_H.

Thus we define the *canonical symplectic form* ω on $Z := V \times V^*$ by

$$\omega((v_1, \alpha_1), (v_2, \alpha_2)) \quad := \quad \alpha_2(v_1) - \alpha_1(v_2). \tag{4.39}$$

So defined, ω is by construction a symplectic form, and so has the normal form given by eq. 4.10.

Given a symplectic vector space (Z, ω), the natural question arises which linear maps $A : Z \to Z$ preserve the normal form given by eq. 4.10. It is straightforward to show that this is equivalent to A preserving the form of Hamilton's equations (for any Hamiltonian); so that these maps A are called *canonical* (or *symplectic*, or *Poisson*). But since (as I announced) this paper does not need details about the theory of canonical transformations, I will not go into details about this. Suffice it to say here the following.

$A : Z \to Z$ is symplectic iff, writing $\tilde{}$ for the transpose (eq. 4.20) and using the second definition eq. 4.35 of $^\sharp$, the following maps (both from Z^* to Z) are equal:

$$A \circ \omega^\sharp \circ \tilde{A} = \omega^\sharp; \tag{4.40}$$

or in matrix notation, with the *matrix* ω given by eq. 4.10, and again writing $\tilde{}$ for the transpose of a matrix

$$A \omega \tilde{A} = \omega. \tag{4.41}$$

(Equivalent formulas are got by taking inverses. We get, respectively: $\tilde{A} \circ \omega^\flat \circ A = \omega^\flat$ and $\tilde{A} \omega A = \omega$.)

The set of all such linear symplectic maps $A : Z \to Z$ form a group, *the symplectic group*, written $\mathrm{Sp}(Z, \omega)$.

To sum up this Subsection: We have, for a vector space V, $\dim(V) = n$, and $Z := V \times V^*$:

(i) the canonical symplectic form $\omega : Z \times Z \to \mathbb{R}$; with normal form given by eq. 4.10;

(ii) the associated linear map $\omega^\flat : Z \to Z^*$; which is an isomorphism, since ω is non-degenerate;

(iii) the associated linear map $\omega^\sharp : Z^* \to Z$; which is an isomorphism, since ω is non-degenerate; and is the inverse of $\omega^\flat$; (cf. eq. 4.35).

We will see shortly that Hamiltonian mechanics takes V to be the tangent space T_q at a point $q \in Q$, so that Z is $T_q \times T_q^*$, i.e. the tangent space to the space Γ of the qs and ps.

5 POISSON BRACKETS AND NOETHER'S THEOREM

We have seen how a single scalar function H on phase space Γ determines the evolution of the system via a combination of partial differentiation (the gradient of H) with the symplectic matrix. We now express these ideas in terms of Poisson brackets.

For our purposes, Poisson brackets will have three main advantages; which will be discussed in the following order in the Subsections below. Poisson brackets:

(i) give a neat expression for the rate of change of any dynamical variable;
(ii) give a version of Noether's theorem which is more simple and powerful (and even easier to prove!) than the Lagrangian version; and
(iii) lead to the generalized Hamiltonian framework mentioned in Section 6.8.

All three advantages arise from the way the Poisson bracket encodes the way that a scalar function determines a (certain kind of) vector field.

5.1 Poisson brackets introduced

The rate of change of any dynamical variable f, taken as a scalar function on phase space $\Gamma, f(q,p) \in \mathbb{R}$, is given (with summation convention) by

$$\frac{df}{dt} = \dot{q}^i \frac{\partial f}{\partial q^i} + \dot{p}_i \frac{\partial f}{\partial p_i}. \tag{5.1}$$

(If f is time-dependent, $f : (q,p,t) \in \Gamma \times \mathbb{R} \mapsto f(q,p,t) \in \mathbb{R}$, the right-hand-side includes a term $\frac{\partial f}{\partial t}$. But on analogy with how our discussion of Lagrangian mechanics imposed scleronomic constraints, a time-independent work-function etc., we here set aside the time-dependent case.) Applying Hamilton's equations, this is

$$\frac{df}{dt} = \frac{\partial H}{\partial p_i}\frac{\partial f}{\partial q^i} - \frac{\partial H}{\partial q^i}\frac{\partial f}{\partial p_i}. \tag{5.2}$$

This suggests that we define the Poisson bracket of any two such functions $f(q,p), g(q,p)$ by

$$\{f,g\} := \frac{\partial f}{\partial q^i}\frac{\partial g}{\partial p_i} - \frac{\partial f}{\partial p_i}\frac{\partial g}{\partial q^i}; \tag{5.3}$$

so that the rate of change of f is given by

$$\frac{df}{dt} = \{f,H\}. \tag{5.4}$$

In terms of the $2n$ coordinates ξ^α (eq. 4.6) and the matrix elements $\omega^{\alpha\beta}$ of ω (eq. 4.13), we can write eq. 5.2 as

$$\frac{df}{dt} = (\partial_\alpha f)\dot{\xi}^\alpha = (\partial_\alpha f)\omega^{\alpha\beta}(\partial_\beta H); \tag{5.5}$$

and so we can define the Poisson bracket by

$$\{f,g\} := (\partial_\alpha f)\omega^{\alpha\beta}(\partial_\beta g) \equiv \frac{\partial f}{\partial \xi^\alpha}\omega^{\alpha\beta}\frac{\partial g}{\partial \xi^\beta}. \tag{5.6}$$

In matrix notation: writing the naive gradients of f and of g as column vectors ∇f and ∇g, and writing $\tilde{}$ for transpose, we have at any point $z = (q,p) \in \Gamma$:

$$\{f,g\}(z) = \tilde{\nabla f}(z) \cdot \omega \cdot \nabla g(z). \tag{5.7}$$

With these definitions of the Poisson bracket, we readily infer the following five results. (Later discussion will bring out the significance of some of these; in particular, Section 6.8 will take some of them to jointly define a primitive Poisson bracket for a generalized Hamiltonian mechanics.)

(1) Since the Poisson bracket is antisymmetric, H itself is a constant of the motion:

$$\frac{dH}{dt} = \{H,H\} \equiv 0. \tag{5.8}$$

(2) The Poisson bracket of a product is given by "Leibniz's rule": i.e. for any three functions f,g,h, we have

$$\{f, h\cdot g\} = \{f,h\}\cdot g + h\cdot\{f,g\}. \tag{5.9}$$

(3) Taking the Poisson bracket as itself a dynamical variable, its time-derivative is given by a "Leibniz rule"; i.e. the Poisson bracket behaves like a product:

$$\frac{d}{dt}\{f,g\} = \left\{\frac{df}{dt}, g\right\} + \left\{f, \frac{dg}{dt}\right\}. \tag{5.10}$$

(4) The Jacobi identity (easily deduced from (3)):

$$\{\{f,h\},g\} + \{\{g,f\},h\} + \{\{h,g\},f\} = 0. \tag{5.11}$$

(5) The Poisson brackets for the qs, ps and ξs are:

$$\{\xi^\alpha, \xi^\beta\} = \omega^{\alpha\beta} \quad ; \quad \text{i.e.} \tag{5.12}$$

$$\{q^i, p_j\} = \delta^i_j, \quad \{q^i, q^j\} = \{p_i, p_j\} = 0. \tag{5.13}$$

Eq. 5.13 is very important, both for general theory and for problem-solving. The reason is that preservation of these Poisson brackets, by a smooth transformation of the $2n$ variables $(q,p) \to (Q(q,p), P(q,p))$, is necessary and sufficient for the transformation being canonical. Besides, in this equivalence 'canonical' can be understood both in the usual elementary sense of preserving the form of Hamilton's equations, for any Hamiltonian function, and in the geometric sense of preserving the symplectic form (explained in (5) of Section 4.3.3, and for manifolds in Section 6).

Note here that, as the phrase 'for any Hamiltonian function' brings out, the notion of a canonical transformation is independent of the forces on the system as encoded in the Hamiltonian. That is: the notion is a matter of Γ's geometry—as we will emphasise in Section 6.

But (as I announced in Section 4.1) I will not need to go into many details about canonical transformations, essentially because this paper does not aim to survey the whole of Hamiltonian mechanics, or even all that can be said about reducing problems, e.g. by finding simplifying canonical transformations. It aims only to survey the way that symmetries and conserved quantities effect such reductions. In the rest of this Subsection, I begin describing Poisson brackets' role in this, in particular Noether's theorem. But the description can only be completed once we have the geometric perspective on Hamiltonian mechanics, i.e. in Section 6.5.

5.2 Hamiltonian vector fields

Section 4.3.1 described how the symplectic matrix enabled the scalar function H on Γ to determine a vector field X_H. The previous Subsection showed how the Poisson bracket expressed any dynamical variable's rate of change along X_H. We now bring these ideas together, and generalize.

Recall that a vector X at a point x of a manifold M can be identified with a directional derivative operator at x assigning to each smooth function f defined on a neighbourhood of x its directional derivative along any curve that has X as its tangent vector. Thus recall the Lagrangian definition of the dynamical vector field, eq. 2.8 in Section 2.2. Similarly here: the dynamical vector field $X_H =: D$ is a derivative operator on scalar functions, which can be written in terms the Poisson bracket:

$$D := X_H = \frac{d}{dt} = \dot{q}^i \frac{\partial}{\partial q^i} + \dot{p}_i \frac{\partial}{\partial p_i} = \frac{\partial H}{\partial p_i}\frac{\partial}{\partial q^i} - \frac{\partial H}{\partial q^i}\frac{\partial}{\partial p_i} = \{\cdot, H\}. \quad (5.14)$$

But this point applies to any smooth scalar, f say, on Γ. That is: although we think of H as the energy that determines the real physical evolution, the mathematics is of course the same for such an f. So any such function determines a vector field, X_f say, on Γ that generates what the evolution "would be if f was the Hamiltonian". Thinking of the integral curves as parametrized by s, we have

$$X_f = \frac{d}{ds} = \{\cdot, f\}. \quad (5.15)$$

X_f is called the *Hamiltonian vector field* of (for) f; just as, for the physical Hamiltonian, $f \equiv H$, Section 4.3.1 called X_H '*the* Hamiltonian vector field'.

The notion of a Hamiltonian vector field will be crucial for what follows, not least for Noether's theorem in the very next Subsection. For the moment, we just make two remarks which we will need later.

So every scalar f determines a Hamiltonian vector field X_f. But note that the converse is false: not every vector field X on Γ is the Hamiltonian vector field of

some scalar. For a vector field (equations of motion) X, with components X^α in the coordinates ξ^α defined by eq. 4.6

$$\dot{\xi}^\alpha = X^\alpha(\xi), \tag{5.16}$$

there need be no scalar $H : \Gamma \to \mathbb{R}$ such that, as required by eq. 4.13,

$$X^\alpha = \omega^{\alpha\beta}\partial_\beta H. \tag{5.17}$$

This is the same point as in (ii) of Section 4.2.3: that Hamilton's equations have the special feature that all the right hand sides are, up to a sign, partial derivatives of a *single* function H—a feature that underpins the possibility of expressing the equations of motion by variational principles.

We also need to note under what condition is a vector field X Hamiltonian; (this will bear on Noether's theorem). The answer is: X is locally Hamiltonian, i.e. there is locally a scalar f such that $X = X_f$, iff X generates a one-parameter family of canonical transformations. We will give a modern geometric proof of this in Section 6.5. For the moment, we only need to note, as at the end of Section 5.1, that here 'canonical transformation' can be understood in the usual elementary sense as a transformation of Γ that preserves the form of Hamilton's equations (for any Hamiltonian); or equivalently, as preserving the Poisson bracket; or equivalently, as preserving the symplectic form (to be defined for manifolds, in Section 6).

5.3 Noether's theorem

5.3.1 An apparent "one-liner", and three claims In the Hamiltonian framework, the core of the proof of Noether's theorem is very simple; as follows. The Poisson bracket is obviously antisymmetric. So for any scalar functions f and H, we have

$$X_f(H) \equiv \frac{dH}{ds} \equiv \{H,f\} = 0 \quad \text{iff} \quad 0 = \{f,H\} = X_H(f) \equiv D(f). \tag{5.18}$$

In words: H is constant under the flow of the vector field X_f (i.e. under what the evolution would be if f was the Hamiltonian) iff f is constant under the dynamical flow $X_H \equiv D$.

This "one-liner" is the Hamiltonian version of Noether's theorem! There are three claims here. The first two relate back to the Lagrangian version of the theorem. The third is about the definition of a (continuous) symmetry for a Hamiltonian system, and so about how we should formulate the Hamiltonian version of Noether's theorem. I will state all three claims, but in this Subsection justify only the first two. For it will be convenient to postpone the third till after we have introduced some modern geometry (Section 6.5).

First, for eq. 5.18 to deserve the name 'Noether's theorem', I need to show that it encompasses Section 3's Lagrangian version of Noether's theorem (despite the trivial proof!).

Second, in order to justify my claim that the Hamiltonian version of Noether's theorem is more powerful than the Lagrangian version, I need to show that eq. 5.18 says more than that version, i.e. that it covers more symmetries.

To state the third claim, note first that we expect a Hamiltonian version of Noether's theorem to say something like: *to every continuous symmetry of a Hamiltonian system, there corresponds a conserved quantity*. Here, we expect a 'continuous symmetry' to be defined by a vector field on Γ (or by its flow). Indeed, a *symmetry* of a Hamiltonian system is usually defined as a transformation of Γ that:

(1) is canonical; (a condition independent of the forces on the system as encoded in the Hamiltonian: a matter of Γ's intrinsic geometry); and also
(2) preserves the Hamiltonian function; (a condition obviously dependent on the Hamiltonian).

Accordingly, a *continuous symmetry* is defined as a vector field on Γ that generates a one-parameter family of such transformations; (or as such a field's flow, i.e. as the family itself).

But with this definition of 'continuous symmetry' (of a Hamiltonian system), eq. 5.18 seems to suffer from two *lacunae*, if taken to express Noether's theorem, that to every continuous symmetry there corresponds a conserved quantity. Agreed, the rightward implication of eq. 5.18 provides, for a vector field X_f with property (2), the conserved quantity f. But there seem to be two *lacunae*:

(a) eq. 5.18 is silent about whether X_f has property (1), i.e. generates canonical transformations.
(b) eq. 5.18 considers only Hamiltonian vector fields, i.e. vector fields X induced by some f, $X = X_f$. But as noted at the end of Section 5.2, there are countless vector fields on Γ that are not Hamiltonian. If such a field could be a continuous symmetry, eq. 5.18's rightward implication would fall short of saying that to *every* continuous symmetry, there corresponds a conserved quantity.

So the third claim I need is that these *lacunae* are illusory. In fact, a single result will deal with both (a) and (b). Namely, it will suffice to show that a vector field X on Γ has property (1), i.e. generates canonical transformations, iff it is Hamiltonian, i.e. induced by some f, $X = X_f$. But I postpone showing this till we have more modern geometry in hand; cf. Section 6.5.

5.3.2 The relation to the Lagrangian version On the other hand, we can establish the first two claims with the elementary apparatus so far developed. I will concentrate on justifying the first claim; that will also make the second claim clear.

For the first claim, we need to show that:

(i) to any variational symmetry of the Lagrangian L, i.e. a vector field X on Q obeying eq. 3.6, there corresponds a vector field X_f on Γ for which $X_f(H) = 0$; and
(ii) the correspondence in (i) is such that the scalar f can be taken to be (the Hamiltonian version of) the momentum p_X conjugate to X, defined by eq. 3.12 (or geometrically, by 3.31).

It will be clearest to proceed in two stages.

(A) First, I will show (i) and (ii).

(B) Then I will discuss how (A) relates to the usual definition of a symmetry of a Hamiltonian system.

(A) The easiest way to show (i) and (ii) is to use the fact discussed after eq. 3.20, that every variational symmetry X arises, around a point where it is non-zero, from a cyclic coordinate in some local system of coordinates. (Recall that this follows from the basic "rectification" theorem securing the local existence and uniqueness of solutions of ordinary differential equations.) That is, there is some coordinate system (q) on some open subset of X's domain of definition on Q such that

(a) X being a variational symmetry is equivalent to q^n being cyclic, i.e. $\frac{\partial L}{\partial \dot{q}^n} = 0$;

(b) the momentum p_X, which the Lagrangian theorem says is conserved, is the elementary generalized momentum $p_n := \frac{\partial L}{\partial \dot{q}^n}$.

So suppose given a variational symmetry X, and a coordinate system (q) satisfying (a)–(b). Now we recall that the Legendre transformation, i.e. the transition between Lagrangian and Hamiltonian frameworks, does not "involve the dependence on the qs". More precisely, we recall eq. 4.8, $\frac{\partial H}{\partial q^n} = -\frac{\partial L}{\partial q^n}$. Now consider $p_n : \Gamma \to \mathbb{R}$. This p_n will do as the function f required in (i) and (ii) above, since

$$X_{p_n}(H) \equiv \{H, p_n\} = \frac{\partial H}{\partial q^n} = -\frac{\partial L}{\partial q^n} = 0. \tag{5.19}$$

Applying eq. 5.18 to eq. 5.19, we deduce that p_n, i.e. the p_X of the Lagrangian theorem, is conserved.

(Hence my remark after eq. 4.8, that the elementary result that p_n is conserved iff q^n is cyclic, underpins the Hamiltonian version of Noether's theorem; just as the corresponding Lagrangian result underpins the Lagrangian version of Noether's theorem: cf. discussion after eq. 3.20.)

(B): I agree that this simple proof seems *suspiciously* simple. Besides, the suspicion grows when you notice that my argument in (A) has not used a definition of a symmetry, in particular a continuous symmetry, of a Hamiltonian system (contrast Section 3.2). As discussed in Section 5.3.1, we expect a Hamiltonian version of Noether's theorem to say 'to every continuous symmetry of a Hamiltonian system there corresponds a conserved quantity'; where a continuous symmetry is a vector field that (1) generates canonical transformations and (2) preserves the Hamiltonian. So the argument in (A) is suspicious since, although eq. 5.19, or the left hand side of eq. 5.18, obviously expresses property (2), i.e. preserving the Hamiltonian, the argument in (A) seems to nowhere use property (1), i.e. the symmetry generating canonical transformations.

But in fact, all is well. The reason why lies in the fact mentioned in (i), (a) of Section 4.1: that every point transformation (together with its lift to TQ) defines a corresponding canonical transformation on T^*Q. That is to say: property (1) is secured by the fact that the Lagrangian Noether's theorem of Section 3 is restricted to symmetries induced by point transformations.

In other words, in terms of the vector field (variational symmetry) X given us by (a) in (A) above: one can check that X defines a vector field on Γ (equivalently: a one-parameter family of transformations on Γ) that is canonical, i.e. preserves Hamilton's equations or equivalently the symplectic form. Indeed, one can easily check that, once we rectify the Lagrangian variational symmetry X, so that it generates the rectified one-parameter family of point transformations: $q_i = \text{const}, i \neq n; q_n \mapsto q_n + \epsilon$, the vector field that X defines on Γ is precisely the field X_{p_n} chosen above.[15]

Finally, the discussion in (B) also vindicates the second claim in Section 5.3.1: that the Hamiltonian version of Noether's theorem, eq. 5.18, *says more* than the Lagrangian version, i.e. covers more symmetries. This follows from the fact (announced in (i) (b) of Section 4.1) that there are canonical transformations *not* induced by a point transformation (together with its lift).

In elementary discussions, this is often expressed in terms of canonical transformations being allowed to "mix" the qs and ps. But a more precise, and geometric, statement is the result announced at the end of Section 5.2 (whose proof is postponed to Section 6.5): that the condition for a vector field on Γ to generate a one-parameter family of canonical transformations is merely that it be a Hamiltonian vector field. That is: for *any* scalar $f : \Gamma \to \mathbb{R}$, the vector field X_f generates such a family.

In this sense, canonical transformations are two a penny (also known as: a dime a dozen!). So it is little wonder that most discussions emphasise the *other* condition, i.e. property (2): that X_f preserve the Hamiltonian, $X_f(H) = 0$. Only very special fs will satisfy $X_f(H) = 0$; and if we are given H (in certain coordinates q, p), it can be very hard to find (the coordinate expression of) such an f.

Indeed, when Jacobi first propounded the theory of canonical transformations, in his *Lectures on Dynamics* (1842), he was of course aware of this. Accordingly, he pointed out that in theoretical mechanics, it was often more fruitful to first consider an f (equivalently: a canonical transformation), and then cast about for a Hamiltonian that it preserved. He wrote: 'The main difficulty in integrating a given differential equation lies in introducing convenient variables, which there is no rule for finding. Therefore we must travel the reverse path and after finding some notable substitution, look for problems to which it can be successfully applied'; (quoted in Arnold (1989, p. 266)). The fact that Jacobi solved many previously intractable problems bears witness to the power of this strategy, and of his theory of canonical transformations.

We can sum up this Subsection in two comments:

(1) In Hamiltonian mechanics, Noether's theorem is a biconditional, an 'iff' statement. Not only does a Hamiltonian symmetry—i.e. a vector field X on Γ that generates canonical transformations (equivalently: preserves the symplectic form, or the Poisson bracket) and preserves the Hamiltonian, $X(H) = 0$—provide a constant of the motion. Also, given a constant of the motion $f : \Gamma \to \mathbb{R}$, there is a symmetry of the Hamiltonian, viz. the vector field X_f. (Or if one prefers the integral notion of symmetry: the flow of X_f). This converse implication, from constant to symmetry, contrasts with the Lagrangian framework; cf. the end of Section 3.4.1.

(2) In elementary Hamiltonian mechanics, Noether's theorem has a very simple one-line proof, viz. eq. 5.18.

Later, we will return to Noether's theorem. Section 6.5 will justify the third claim of Section 5.3.1, by showing that a vector field generates a one-parameter family of canonical transformations iff it is a Hamiltonian vector field. Meanwhile, we end Section 5 with a comment about "iterating" Noether's theorem, and the distinction between such an iteration and the idea of complete integrability.

5.4 Glimpsing the "complete solution"

Suppose we "iterate" Noether's theorem. That is: suppose there are several (continuous) symmetries of the Hamiltonian and so several constants of the motion. Each will confine the system's time-evolution to a $(2n-1)$-dimensional hypersurface of Γ. In general, the intersection of k such surfaces will be a hypersurface of dimension $2n-k$ (i.e. of co-dimension k); to which the motion is therefore confined. The theory of symplectic reduction (Butterfield 2006) describes how to do a "quotiented dynamics" in this general situation. Here, I just remark on one aspect; which will *not* be developed in the sequel.

Locally, the rectification theorem secures, for any system, not just several constants of the motion, but "all you could ask for". Applying the theorem (eq. 3.21 and 3.22) to the Hamiltonian vector field X_H on Γ, we infer that locally there are coordinates ξ^α (maybe very hard to find!) in which X_H has $2n-1$ components that vanish throughout the neighbourhood, while the other component is 1:

$$X_H^\alpha = 0 \text{ for } \alpha = 1, 2, \ldots, 2n-1; \; X_H^{2n} = 1. \tag{5.20}$$

So the coordinates $\xi^\alpha, \alpha = 1, \ldots, 2n-1$, form $2n-1$ constants of the motion. They are functionally independent, and all other constants of the motion are functions of them; (cf. point (ii) after eq. 3.22). So the motion is confined to the one-dimensional intersection of the $2n-1$ hypersurfaces, each of co-dimension 1. That is to say, it is confined to the curve given by: $\xi^\alpha = \text{const}, \alpha = 1, \ldots, 2n-1, \xi^{2n} = t$.

To this, Noether's theorem eq. 5.18 adds the physical idea that each such constant of the motion defines a vector field X_{ξ^α} that generates a symmetry of the Hamiltonian:

$$X_{\xi^\alpha}(H) = 0, \text{ for } \alpha = 1, 2, \ldots, 2n-1. \tag{5.21}$$

In this local sense, the "complete solution" of any Hamiltonian system lies in the local constants of the motion, or equivalently the local symmetries of its Hamiltonian H.

To sum up: locally, any Hamiltonian system is "completely integrable". But the scare-quotes here are a reminder that these phrases are usually used with other, stronger, meanings: either that there are $2n-1$ *global* constants of the motion or that the system is completely integrable in the sense of Liouville's theorem.

6 A GEOMETRICAL PERSPECTIVE

In this final Section, we develop the modern geometric description of Hamiltonian mechanics. We will build especially on Sections 4.3; one main aim will of course be to complete the discussion of Noether's theorem, begun in Section 5.3.

There will be eight Subsections. First, we introduce the cotangent bundle T^*Q. Then we collect what we will need about forms. Then we can show that any cotangent bundle is a symplectic manifold. This enables us to formulate Hamilton's equations geometrically; and to complete the discussion of Noether's theorem. Then we report Darboux's theorem, and its relation to reduction of problems. Then we return to the Lagrangian framework, by sketching the geometric formulation of the Legendre transformation. Finally, we "glimpse the landscape ahead" by mentioning the more general framework for Hamiltonian mechanics that uses Poisson manifolds.

6.1 Canonical momenta are one-forms: Γ *as* T^*Q

So far we have treated the phase space Γ informally: saying just that it is a $2n$-dimensional space coordinatized by the qs, a smooth coordinate system on the configuration manifold Q, and the ps, which are canonical momenta $\frac{\partial L}{\partial \dot{q}^i}$. But we also saw in (3) of Section 2.2.2 that at each point $q \in Q$, the p_i transform as a 1-form (eq. 2.12). Accordingly we now take the physical state of the system to be a point in the cotangent bundle T^*Q, the $2n$-dimensional manifold whose points are pairs (q, p) with $q \in Q, p \in T^*_q$.

I stress that from now on, the symbol p has a (fruitful!) ambiguity, between "dynamics" and "kinematics/geometry". For p represents both:

(A) the conjugate momentum $\frac{\partial L}{\partial \dot{q}}$, which of course depends on the choice of L; and

(B) a point in a fibre T^*_q of the cotangent bundle T^*Q (i.e. a 1-form or covector); or relatedly: the components p_i of such a 1-form: notions that are independent of any choice of a Lagrangian or Hamiltonian.

In more detail:

(A) Recall that in the Lagrangian framework, the basic equations (eq. 2.1, or Newton's second law!) being second-order in time prompts us to take the initial q and $\dot{q}$ as chosen independently, with L (encoding the forces on the system) then determining the evolution (the Lagrangian dynamical vector field D)—and so also determining the actual "realized" value of $\dot{q}$ at other times as a function of q, and so ultimately, of t. Similarly here: Newton's second law being second-order in time prompts us to take the initial q and p as independent, with H (encoding the forces on the system) then determining the evolution (the Hamiltonian dynamical vector field D)—and so also determining the actual value of p at other times as a function of q, and so ultimately, of t. Besides, by passing via the Legendre transformation back to the Lagrangian framework, one can check that the later actual value of p is determined to equal $\frac{\partial L}{\partial \dot{q}}$.

(B) But p also represents any 1-form (so that p_i represents the 1-form's coordinates). Here, we need to recall three points:—

(i) A local coordinate system (a chart) on Q defines a basis in the tangent space T_q at any point q in the chart's domain. As usual, I write the chart's coordinate functions as q^i. So I shall temporarily denote the chart by $[q]$, so that there are coordinate functions $q^i : \mathrm{dom}([q]) \to \mathbb{R}$. I write elements of the coordinate basis as usual, as $\frac{\partial}{\partial q^i}$.

(ii) The chart $[q]$ thereby also defines a dual basis dq^i in the cotangent space T^*_q at any $q \in \mathrm{dom}([q])$.

(Here I recall, *en passant*, that the isomorphism at each q between T_q and T^*_q, that maps the basis element $\frac{\partial}{\partial q^i} \in T_q$ to the one-form dq^i in the dual basis, is basis-*dependent*. A different basis $\frac{\partial}{\partial q'^i}$ would give a different isomorphism. Cf. the discussion in (1) of Section 4.3.3.)

(iii) Putting (i) and (ii) together: the chart $[q]$ thereby also induces a local coordinate system on a neighbourhood of the cotangent bundle around any point $(q,p) \in T^*Q$ with $q \in \mathrm{dom}([q])$ and $p \in T^*_q$.

Putting (i)–(iii) together: the coordinates of any point (q,p) in T^*Q in such a coordinate system are usually also written as (q,p). That is: p is used for the components of *any* 1-form, in the basis dq^i dual to a coordinate basis $\frac{\partial}{\partial q^i}$. So, similarly to (i) above: I will write this induced chart on T^*Q as $[q,p]$.

(C) Taken together, points (A) and (B) prompt a question:

> Why should an evolution from an arbitrary initial state $\in T^*Q$ have the property that:
> *if* we choose to express
> (i) its configuration, q_0 say, in terms of an arbitrary initial coordinate system $[q]$ on Q, and
> (ii) its momenta $\frac{\partial L}{\partial \dot{q}}$ in terms of the basis dq dual to the coordinate basis $\frac{\partial}{\partial q}$ at q_0:
> *then*
> the states at a *later* time t have *their* momenta—which the Lagrangian framework tells us must be $\frac{\partial L}{\partial \dot{q}}$ (cf. (A))—equal to their components in the dual basis to the *later* coordinate basis, i.e. the coordinate basis $\frac{\partial}{\partial q}$ at the later configuration q_t?
>
> In short: why should the state's components in the dual basis of any coordinate basis continue to be equal, as dynamical evolution goes on, to the values of canonical momenta i.e. $\frac{\partial L}{\partial \dot{q}}$?

A good question. The short answer lies in combining Hamilton's equations for the time-derivative of the p_i (eq. 4.5) with Lagrange's equations, and with the fact that the partial derivatives with respect to q^i of the Hamiltonian and Lagrangian, H and

L, are negatives of each other (eq. 4.8). Thus we have:

$$\dot{p}_i = -\frac{\partial H}{\partial q^i} = \frac{\partial L}{\partial q^i} = \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}^i}\right). \tag{6.1}$$

From this it is clear that for any coordinate system, if at t_0, p_i is chosen to equal $\frac{\partial L}{\partial \dot{q}^i}$, then this will be so at later times. For eq. 6.1 forces their time-derivatives to be equal—and so also, their later values must be equal.

So much for the short answer. We will also get more insight into the relations between the Lagrangian and Hamiltonian frameworks in

(i) the fact, expounded in Section 6.3 below, that any cotangent bundle has a natural symplectic structure, independent of the specification of any Lagrangian or Hamiltonian function; and

(ii) some further details about the Legendre transformation, which is further discussed in Section 6.7.

6.2 Forms, wedge-products and exterior derivatives

As I said at the end of Section 4.3.2, this paper can largely avoid the theory of forms. For what follows (especially Section 6.5), I need to recall only:

(i) the idea of forms of various degrees, together comprising the exterior algebra, and equipped with operations of wedge-product and contraction (Section 6.2.1);

(ii) the ideas of differential forms, the exterior derivative, and of exact and closed forms (Section 6.2.2).

6.2.1 The exterior algebra; wedge-products and contractions We begin by recalling some ideas of Sections 4.3.2 and 4.3.3. Let us again begin with the simplest possible case, $\mathbb{R}^2$, considered as a vector space: not as a manifold with a copy of itself as tangent space at each point.

If α, β are covectors, i.e. elements of $(\mathbb{R}^2)^*$, we define their *wedge-product*, an antisymmetric bilinear form on $\mathbb{R}^2$, by

$$\alpha \wedge \beta : (v, w) \in \mathbb{R}^2 \times \mathbb{R}^2 \mapsto (\alpha(v))(\beta(w)) - (\alpha(w))(\beta(v)) \in \mathbb{R}. \tag{6.2}$$

Let us write the standard basis elements of $\mathbb{R}^2$ as $\frac{\partial}{\partial q}$ and $\frac{\partial}{\partial p}$, with elements of $\mathbb{R}^2$ having components (q, p) in this basis; and let us write the elements of the dual basis as dq, dp. Recalling the definition of the area form A, eq. 4.16, we deduce that A is $dq \wedge dp$.

Similarly for $\mathbb{R}^{2n}$. Recall that the symplectic matrix defines an antisymmetric bilinear form on $\mathbb{R}^{2n}$ by eq. 4.18. The value on a pair $(q, p) \equiv (q^1, \ldots, q^n; p_1, \ldots, p_n), (q', p') \equiv (q'^1, \ldots, q'^n; p'_1, \ldots, p'_n)$ is the sum of the signed areas of the n parallelograms formed by the projections of the vectors $(q, p), (q', p')$ onto the n pairs of coordinate planes. This is a sum of n wedge-products. That is to say: if we write the standard basis elements as $\frac{\partial}{\partial q^i}$ and $\frac{\partial}{\partial p_i}$, this form is $\omega := \Sigma_i \, dq^i \wedge dp_i$.

It has the action on $\mathbb{R}^n \times \mathbb{R}^n$:

$$\left(q^i \frac{\partial}{\partial q^i} + p_i \frac{\partial}{\partial p_i}, q'^i \frac{\partial}{\partial q^i} + p'_i \frac{\partial}{\partial p_i} \right) \mapsto \sum_{i=1}^{n} q^i p'_i - q'^i p_i. \tag{6.3}$$

In general, if V, W are two (real finite-dimensional) vector spaces, we define: $L(V, W)$ to be the vector space of linear maps from V to W; $L^k(V, W)$ to be the vector space of k-multilinear maps from $V \times V \times \cdots \times V$ (k copies) to W; and $L^k_a(V, W)$ to be the subspace of $L^k(V, W)$ consisting of (wholly) antisymmetric maps.

We then define $\Omega^k(V) := L^k_a(V, \mathbb{R})$ for $k = 1, 2, \ldots, \dim(V)$, so that $\Omega^1(V) = V^*$. We also set $\Omega^0(V) := \mathbb{R}$. $\Omega^k(V)$ is called the space of *(exterior) k-forms* on V. If $\dim(V) = n$, then $\dim(\Omega^k(V)) = \binom{n}{k}$.

The wedge-product, as defined above, can be extended to be an operation that defines, for $\alpha \in \Omega^k(V), \beta \in \Omega^l(V)$, an element $\alpha \wedge \beta \in \Omega^{k+l}(V)$. We can skip the details: suffice it to say that the idea is to take tensor products as in (3) of Section 4.3.3, and anti-symmetrize.

But to complete our discussion of Noether's theorem (in Section 6.5), we will need the definition of the *contraction*, (also known as: *interior product*), of a k-form $\alpha \in \Omega^k(V)$ with a vector $v \in V$. We shall write this as $\mathbf{i}_v\alpha$. (It is also written with a hook notation.) We define the contraction $\mathbf{i}_v\alpha$ to be the $(k-1)$-form given by:

$$\mathbf{i}_v\alpha(v_2, \ldots, v_k) := \alpha(v, v_2, \ldots, v_k). \tag{6.4}$$

It follows, for example, that contraction distributes over the wedge-product *modulo* a sign, in the following sense. If α is a k-form, and β a 1-form, then

$$\mathbf{i}_v(\alpha \wedge \beta) = (\mathbf{i}_v\alpha) \wedge \beta + (-1)^k \alpha \wedge (\mathbf{i}_v\beta) \ . \tag{6.5}$$

The direct sum of the vector spaces $\Omega^k(V), k = 0, 1, 2, \ldots, \dim(V) =: n$, has dimension 2^n. When this direct sum is considered as equipped with the wedge-product $\wedge$ and contraction $\mathbf{i}$, it is called the *exterior algebra* of V, written $\Omega(V)$.

6.2.2 Differential forms; the exterior derivative; the Poincaré Lemma We extend the discussion given in Section 6.2.1 to a manifold M of dimension n, taking all the tangent spaces T_x at $x \in M$ as copies of the vector space V, and requiring fields of forms to be suitably smooth.

We begin by saying that a (smooth) scalar function $f : M \to \mathbb{R}$ is a 0-form field. Its *differential* or *gradient*, df, as defined by its action on all vector fields X, viz. mapping them to f's directional derivative along X

$$df(X) := X(f) \tag{6.6}$$

is a 1-form (covector) field, called a *differential 1-form*.

The set $\mathcal{F}(M)$ of all smooth scalar functions forms an (infinite-dimensional) vector space, indeed a ring, under pointwise operations. We write the set of vector fields on M as $\mathcal{X}(M)$, or as $\mathcal{T}_0^1(M)$; and the set of covector fields, i.e. differential 1-forms, on M as $\mathcal{X}^*(M)$, or as $\mathcal{T}_1^0(M)$. (So superscripts indicate the contravariant order, and subscripts the covariant order.)

Accordingly, we define: $\Omega^0(M) := \mathcal{F}(M)$; $\Omega^1(M) = \mathcal{T}_1^0(M)$; and so on. In short: $\Omega^k(M)$ is the set of smooth fields of exterior k-forms on the tangent spaces of M.

The wedge-product, as defined in Section 6.2.1, can be extended to the various $\Omega^k(M)$. We form the direct sum of the (infinite-dimensional) vector spaces $\Omega^k(M), k = 0, 1, 2, \ldots, \dim(V) =: n$, and consider it as equipped with this extended wedge-product. We call it the *algebra of exterior differential forms* on M, written $\Omega(M)$.

Similarly, contraction, as defined in Section 6.2.1, can be extended to $\Omega(M)$. On analogy with eq. 6.4, we define, for α a k-form field on M, and X a vector field on M, the contraction $\mathbf{i}_X\alpha$ to be the $(k-1)$-form given, at each point $x \in M$, by:

$$\mathbf{i}_X\alpha(x) : (v_2, \ldots, v_k) \mapsto \alpha(x)(X(x), v_2, \ldots, v_k) \in \mathbb{R} \ . \tag{6.7}$$

The *exterior derivative* is a differential operator on $\Omega(M)$ that maps a k-form field to a $(k+1)$-form field. In particular, it maps a scalar f to its differential (gradient) df. Indeed, it is the *unique* map from the k-form fields to the $(k+1)$-form fields $(k = 1, 2, \ldots, n)$ that generalizes the elementary notion of gradient $f \mapsto df$, subject to certain natural conditions.

To be precise: one can show that there is a unique family of maps $d^k : \Omega^k(M) \to \Omega^{k+1}(M)$, all of which, for simplicity, we write as $\mathbf{d}$, such that:

(a) If $f \in \mathcal{F}(M)$, $\mathbf{d}(f) = df$.

(b) $\mathbf{d}$ is $\mathbb{R}$-linear; and distributes across the wedge-product, *modulo* a sign. That is: for $\alpha \in \Omega^k(M), \beta \in \Omega^l(M)$, $\mathbf{d}(\alpha \wedge \beta) = (\mathbf{d}\alpha) \wedge \beta + (-1)^k \alpha \wedge (\mathbf{d}\beta)$. (Cf. eq. 6.5.)

(c) $\mathbf{d}^2 := \mathbf{d} \circ \mathbf{d} \equiv 0$; i.e. for all $\alpha \in \Omega^k(M)$ $d^{k+1} \circ d^k(\alpha) \equiv 0$. (This condition looks strong, but is in fact natural. For its motivation, it must here suffice to say that it generalizes the fact in elementary vector calculus, that the curl of any gradient is zero: $\nabla \wedge (\nabla f) \equiv 0$.)

(d) $\mathbf{d}$ is a *local operator*; i.e. for any $x \in M$ and any k-form α, $\mathbf{d}\alpha(x)$ depends only on α's restriction to any open neighbourhood of x; more precisely, we define for any open set U of M, the vector space $\Omega^k(U)$ of k-form fields on U, and then require that

$$\mathbf{d}(\alpha \mid_U) = (\mathbf{d}\alpha) \mid_U. \tag{6.8}$$

To express $\mathbf{d}$ in terms of coordinates: if $\alpha \in \Omega^k(M)$, i.e. α is a k-form on M, given in coordinates by

$$\alpha = \alpha_{i_1 \ldots i_k} \, dx^{i_1} \wedge \cdots \wedge dx^{i_k} \quad (\text{sum on } \ i_1 < i_2 < \cdots < i_k), \tag{6.9}$$

then one proves that the exterior derivative is

$$\mathbf{d}\alpha \;=\; \frac{\partial \alpha_{i_1 \ldots i_k}}{\partial x^j}\, dx^j \wedge dx^{i_1} \wedge \cdots \wedge dx^{i_k} \quad (\text{sum on all } j \text{ and } i_1, \cdots < i_k), \tag{6.10}$$

We define $\alpha \in \Omega^k(M)$ to be:

exact if there is a $\beta \in \Omega^{k-1}(M)$ such that $\alpha = \mathbf{d}\beta$; (cf. the elementary definition of an exact differential);

closed if $\mathbf{d}\alpha = 0$.

It is immediate from condition (c) above, $\mathbf{d}^2 = 0$, that every exact form is closed. The converse is "locally true". This important result is the *Poincaré Lemma*; (and we will use it in Section 6.5's closing discussion of Noether's theorem).

To be precise: for any open set U of M, we define (as in condition (d) above) the vector space $\Omega^k(U)$ of k-form fields on U. Then the *Poincaré Lemma* states that if $\alpha \in \Omega^k(M)$ is closed, then at every $x \in M$ there is a neighbourhood U such that $\alpha\,|_U \in \Omega^k(U)$ is exact.

We will also need (again, for Section 6.5's discussion of Noether's theorem) a useful formula relating the Lie derivative, contraction and the exterior derivative. Namely: *Cartan's magic formula*, which says that if X is a vector field and α a k-form on a manifold M, then the Lie derivative of α with respect to X (i.e. along the flow of X) is

$$\mathcal{L}_X \alpha = \mathbf{d}\mathbf{i}_X \alpha + \mathbf{i}_X \mathbf{d}\alpha. \tag{6.11}$$

This is proved by straightforward calculation.

6.3 Symplectic manifolds; the cotangent bundle as a symplectic manifold

Any cotangent bundle T^*Q has a natural *symplectic structure*, which is the geometric structure on manifolds corresponding to the symplectic matrix ω introduced by eq. 4.10, and to the symplectic forms on vector spaces defined at the end of Section 4.3.3. (Here 'natural' means intrinsic, and in particular, independent of a choice of coordinates or bases.) It is this structure that enables a scalar function to determine a dynamics. That is: the symplectic structure implies that any scalar function $H : T^*Q \to \mathbb{R}$ defines a vector field X_H on T^*Q.

I first describe this structure (Section 6.3.1), and then show that any cotangent bundle has it (Section 6.3.2). Later subsections will develop the consequences.

6.3.1 Symplectic manifolds *A symplectic structure* or *symplectic form* on a manifold M is defined to be a differential 2-form ω on M that is closed (i.e. $\mathbf{d}\omega = 0$) and non-degenerate. That is: for any $x \in M$, and any two tangent vectors at x, $\sigma, \tau \in T_x$:

$$\mathbf{d}\omega = 0 \quad \text{and} \quad \forall\, \tau \neq 0,\ \exists \sigma : \quad \omega(\tau, \sigma) \neq 0. \tag{6.12}$$

Such a pair (M, ω) is called a *symplectic manifold*.

There is a rich theory of symplectic manifolds; but we shall only need a small fragment of it, building on our discussion in Section 4.3.3. (In particular, the fact that we mostly avoid the theory of canonical transformations means we will not need the theory of Lagrangian sub-manifolds.)

First, it follows from the non-degeneracy of ω that M is even-dimensional; (cf. eq. 4.38).

It also follows that at any $x \in M$, there is a basis-independent isomorphism $\omega^\flat$ from the tangent space T_x to its dual T_x^*. We saw this in (2) and (4) of Section 4.3.3, especially eq. 4.23. Namely: for any $x \in M$ and $\tau \in T_x$, the value of the 1-form $\omega^\flat(\tau) \in T_x^*$ is defined by

$$\omega^\flat(\tau)(\sigma) := \omega(\sigma, \tau) \quad \forall \sigma \in T_x. \tag{6.13}$$

Here we return to the main idea emphasised already in Section 4.3.1: that symplectic structure enables a covector field, i.e. a differential one-form, to determine a vector field. Thus for any function $H : M \to \mathbb{R}$, so that dH is a differential 1-form on M, the inverse of $\omega^\flat$ (which we might write as $\omega^\sharp$), carries dH to a vector field on M, written X_H. Cf. eq. 4.14.

So far, we have noted some implications of ω being non-degenerate. The other part of the definition of a symplectic form (for a manifold), viz. ω being closed, $\mathbf{d}\omega = 0$, is also important. We shall see in Section 6.5 that it implies that a vector field X on a symplectic manifold M preserves the symplectic form ω (i.e. in more physical jargon: generates (a one-parameter family of) canonical transformations) iff X is Hamiltonian in the sense of Section 5.2; i.e. there is a scalar function f such that $X = X_f \equiv \omega^\sharp(df)$. Or in terms of the Poisson bracket, with $\cdot$ representing the argument place for a scalar function: $X(\cdot) = X_f(\cdot) \equiv \{\cdot, f\}$.

So much by way of introducing symplectic manifolds. I turn to showing that any cotangent bundle T^*Q is such a manifold.

6.3.2 The cotangent bundle Choose any local coordinates q on Q ($\dim(Q)=n$), and the natural local coordinates q,p thereby induced on T^*Q; (cf. (B) of Section 6.1). We define the 2-form

$$dp \wedge dq := dp_i \wedge dq^i := \Sigma_{i=1}^n dp_i \wedge dq^i. \tag{6.14}$$

To show that eq. 6.14 defines the same 2-form, whatever choice we make of the chart q on Q, it suffices to show that $dp \wedge dq$ is the exterior derivative of a 1-form on T^*Q which is defined naturally (i.e. independently of coordinates or bases) from the derivative (also known as: tangent) map of the projection

$$\pi : (q,p) \in T^*Q \mapsto q \in Q. \tag{6.15}$$

Thus consider a tangent vector τ (not to Q, but) to the cotangent bundle T^*Q at a point $\eta = (q,p) \in T^*Q$, i.e. $q \in Q$ and $p \in T_q^*$. Let us write this as: $\tau \in T_\eta(T^*Q) \equiv T_{(q,p)}(T^*Q)$. The derivative map, $D\pi$ say, of the natural projection π applies to τ:

$$D\pi : \tau \in T_{(q,p)}(T^*Q) \mapsto (D\pi(\tau)) \in T_q. \tag{6.16}$$

Now define a 1-form θ_H on T^*Q by

$$\theta_H : \tau \in T_{(q,p)}(T^*Q) \mapsto p(D\pi(\tau)) \in \mathbb{R}; \tag{6.17}$$

where in this definition of θ_H, p is defined to be the second component of τ's base-point $(q,p) \in T^*Q$; i.e. $\tau \in T_{(q,p)}(T^*Q)$ and $p \in T^*_q$.

This 1-form is called the *canonical 1-form* on T^*Q. It is the "Hamiltonian version" of the 1-form θ_L defined by eq. 2.13; and also there called the 'canonical 1-form'. But Section 6.1's discussion of the "fruitful ambiguity" of the symbol p brings out a contrast. While θ_L as defined by eq. 2.13 clearly depends on L, the definition of θ_H, eq. 6.17, does *not* depend on any function H. θ_H is given just by the cotangent bundle structure. Hence the subscript H here just indicates "Hamiltonian (as against Lagrangian) version", *not* dependence on a function H.

So much by way of a natural definition of a 1-form. One now checks that in any natural local coordinates q, p, θ_H is given by

$$\theta_H = p_i dq^i. \tag{6.18}$$

Finally, we define a 2-form by taking the exterior derivative of θ_H:

$$\mathbf{d}(\theta_H) := \mathbf{d}(p_i dq^i) \equiv dp_i \wedge dq^i. \tag{6.19}$$

where the last equation follows immediately from eq. 6.10. One checks that this 2-form is closed (since $\mathbf{d}^2 = 0$) and non-degenerate. So $(T^*Q, \mathbf{d}(\theta_H))$ is a symplectic manifold.

Referring to eq. 4.18 of Section 4.3, or eq. 4.39 of Section 4.3.3, or eq. 6.3 of Section 6.2, we see that at each point $(q,p) \in T^*Q$, this symplectic form is, upto a sign, our familiar "sum of signed areas"—first seen as induced by the matrix ω of eq. 4.10.

Accordingly, Section 4.3.3's definition of a canonical symplectic form is extended to the present case: $\mathbf{d}(\theta_H)$, or its negative $-\mathbf{d}(\theta_H)$, is called the *canonical symplectic form*, or *canonical 2-form*. (The difference from Section 4.3.3's definition is that on a manifold, the symplectic form is required to be closed.)

(The difference by a sign is of course conventional: it arises from our taking the qs, not the ps, as the first n out of the $2n$ coordinates. For if we had instead taken the ps, the matrix occurring in eq. 4.12 would have been $-\omega \equiv \omega^{-1}$: exactly matching the cotangent bundle's intrinsic 2-form $\mathbf{d}(\theta_H)$.)

We will see, in Section 6.6, a theorem (Darboux's theorem) to the effect that locally, any symplectic manifold "looks like" a cotangent bundle: or in other words, a cotangent bundle is locally a "universal" example of symplectic structure. But first we return, in the next two Subsections, to Hamilton's equations, and Noether's theorem.

6.4 Geometric formulations of Hamilton's equations

We already emphasised in Sections 4.3 and 5 the main geometric idea behind Hamilton's equations: that a gradient, i.e. covector, field dH determines a vector

field X_H. We first saw this determination via the symplectic matrix, in eq. 4.14 of Section 4.3.1, viz.

$$X_H(z) = \omega \nabla H(z); \tag{6.20}$$

and then via the Poisson bracket, in eq. 5.14 of Section 5.2, viz.

$$D := X_H = \frac{d}{dt} = \dot{q}^i \frac{\partial}{\partial q^i} + \dot{p}_i \frac{\partial}{\partial p_i} = \frac{\partial H}{\partial p_i}\frac{\partial}{\partial q^i} - \frac{\partial H}{\partial q^i}\frac{\partial}{\partial p_i} = \{\cdot, H\}\ . \tag{6.21}$$

The symplectic structure and Poisson bracket were related by eq. 5.7, viz.

$$\{f, g\}(z) = \tilde{\nabla} f(z).\omega.\nabla g(z). \tag{6.22}$$

And to this earlier discussion, the last Subsection, Section 6.3, added the identification of the canonical symplectic form of a cotangent bundle, eq. 6.19.

Let us sum up these discussions by giving some geometric formulations of Hamilton's equations at a point $z = (q, p)$ in a cotangent bundle T^*Q. Let us write $\omega^\sharp$ for the (basis-independent) isomorphism from the cotangent space to the tangent space, $T_z^* \to T_z$, induced by $\omega := -\mathbf{d}(\theta_H) = dq^i \wedge dp_i$ (cf. eq. 4.35 and 6.13). Then Hamilton's equations, eq. 4.14 or 6.20, may be written as:

$$\dot{z} = X_H(z) = \omega^\sharp(\mathbf{d}H(z)) = \omega^\sharp(dH(z)). \tag{6.23}$$

Applying $\omega^\flat$, the inverse isomorphism $T_z \to T_z^*$, to both sides, we get

$$\omega^\flat X_H(z) = dH(z). \tag{6.24}$$

In terms of the symplectic form ω at z, this is (cf. eq. 4.23): for all vectors $\tau \in T_z$

$$\omega(X_H(z), \tau) = dH(z) \cdot \tau; \tag{6.25}$$

or in terms of the contraction defined by eq. 6.4, with $\cdot$ marking the argument place of $\tau \in T_z$:

$$\mathbf{i}_{X_H}\omega := \omega(X_H(z), \cdot) = dH(z)(\cdot). \tag{6.26}$$

More briefly, and now for any function f, it is:

$$\mathbf{i}_{X_f}\omega = df. \tag{6.27}$$

Here is a final example. Recall the relation between the Poisson bracket and the directional derivative (or the Lie derivative $\mathcal{L}$) of a function, eq. 5.15 and 6.21: viz.

$$\mathcal{L}_{X_f} g = dg(X_f) = X_f(g) = \{g, f\}. \tag{6.28}$$

Combining this with eq. 6.27, we can reformulate the relation between the symplectic form and Poisson bracket, eq. 6.22, in the form:

$$\{g, f\} = dg(X_f) = \mathbf{i}_{X_f} dg = \mathbf{i}_{X_f}(\mathbf{i}_{X_g}\omega) = \omega(X_g, X_f). \tag{6.29}$$

6.5 *Noether's theorem completed*

The discussion of Noether's theorem in Section 5.3 left unfinished business: to prove that a vector field generates a one-parameter family of canonical transformations iff it is a Hamiltonian vector field (and so justify the third claim of Section 5.3.1). Cartan's magic formula and the Poincaré Lemma, both from Section 6.2, make it easy to prove this, for a vector field on any symplectic manifold (M, ω). ((M, ω) need not be a cotangent bundle.)

We define a vector field X on a symplectic manifold (M, ω) to be *symplectic* (also known as: *canonical*) iff the Lie-derivative along X of the symplectic form vanishes, i.e. $\mathcal{L}_X \omega = 0$.[16]

Since ω is closed, i.e. $\mathbf{d}\omega = 0$, Cartan's magic formula, eq. 6.11, applied to ω becomes

$$\mathcal{L}_X \omega \equiv \mathbf{d}\mathbf{i}_X \omega + \mathbf{i}_X \mathbf{d}\omega = \mathbf{d}\mathbf{i}_X \omega. \tag{6.30}$$

So for X to be symplectic is for $\mathbf{i}_X \omega$ to be closed. But by the Poincaré Lemma, if $\mathbf{i}_X \omega$ is closed, it is locally exact. That is: there locally exists a scalar function $f : M \to \mathbb{R}$ such that

$$\mathbf{i}_X \omega = df \quad \text{i.e. } X = X_f. \tag{6.31}$$

So for X to be symplectic is equivalent to X being *locally Hamiltonian*.

So we can sum up Noether's theorem from a geometric perspective, as follows. We define a *Hamilton system* to be a triple (M, ω, H) where (M, ω) is a symplectic manifold and $H : M \to \mathbb{R}$, i.e. $M \in \mathcal{F}(M)$. We define a (continuous) *symmetry* of a Hamiltonian system to be a vector field X on M that preserves both the symplectic form, $\mathcal{L}_X \omega = 0$, and the Hamiltonian function, $\mathcal{L}_X H = 0$. As we have just seen: for any symmetry so defined, there locally exists an f such that $X = X_f$. So we can apply the "one-liner", eq. 5.18, i.e. the antisymmetry of the Poisson bracket,

$$X_f(H) \equiv \{H, f\} = 0 \quad \text{iff} \quad X_H(f) \equiv \{f, H\} = 0, \tag{6.32}$$

to conclude that f is a first integral (constant of the motion). Thus we have

Noether's theorem for a Hamilton system If X is a symmetry of a Hamiltonian system (M, ω, H), then locally $X = X_f$ and f is a constant of the motion. And conversely: if $f : M \to \mathbb{R}$ is a constant of the motion, then X_f is a symmetry. Besides, this result encompasses the Lagrangian version of the theorem; cf. Sections 3.4 and 5.3.

Example:— For most Hamiltonian systems in euclidean space $\mathbb{R}^3$, spatial translations and rotations are (continuous) symmetries. For example, consider N point-particles interacting by Newtonian gravity. The Hamiltonian is a sum of two

terms, which are each individually invariant under these euclidean motions:

(i) a kinetic energy term K; though I will not go into details, it is in fact defined by the euclidean metric of $\mathbb{R}^3$ (cf. footnote 2 in Section 2.1), and is thereby invariant; and

(ii) a potential energy term V; it depends only on the particles' relative distances, and is thereby invariant.

The corresponding conserved quantities are the total linear and angular momentum.[17]

Finally, an incidental remark which relates to the "rectification theorem", that on any manifold any vector field X can be "straightened out" in a neighbourhood around any point at which X is non-zero, so as to have all but one component vanish and the last component equal to 1; cf. eq. 3.22. Using this theorem, it is easy to see that on any even-dimensional manifold *any* vector field X is locally Hamiltonian, with respect to *some* symplectic form, around a point where X is non-zero. (One defines the symplectic form by Lie-dragging from a surface transverse to X's integral curves.)

6.6 Darboux's theorem, and its role in reduction

Darboux's theorem states that cotangent bundles are, locally, a "universal form" of symplectic manifold. That is: Not only is any symplectic manifold (M, ω) even-dimensional. Also, it "looks locally like" a cotangent bundle, in that around any x in M, there is a local coordinate system $(q^1, \ldots, q^n; p_1, \ldots, p_n)$—where the use of both upper and lower indices is now just conventional, with no meaning about dual bases!—in which:

(i) ω takes the form $dq^i \wedge dp_i$; and so

(ii) the Poisson brackets of the qs and ps take the fundamental form in eq. 5.13.

(The theorem generalizes to the Poisson manifolds mentioned in Section 6.8.)

Besides, the proof of Darboux's theorem yields further information: information which is important for reducing problems. It arises from the beginning of the proof; and will return us to Section 4.2's point that the elementary connection between cyclic coordinates and conserved conjugate momenta underpins the role of symmetries and conserved quantities in reductions on symplectic manifolds.

(In fact, Darboux's theorem also yields two other broad implications about reducing problems; but I will not develop the details here. The second implication concerns the way that a Hamiltonian structure is preserved in the reduced problem. The third implication concerns the requirement that constants of the motion be in involution, i.e. have vanishing Poisson bracket with each other; so it leads to the idea of complete integrability—a topic this paper foreswears.)

Namely, the proof implies that "almost" any scalar function $f \in \mathcal{F}(M)$ can be taken as the first "momentum" coordinate p_1; or as the first configurational coordinate q^1. Here "almost" is not meant in a measure-theoretic sense; it is just that f is subject to a mild restriction, that $df \neq 0$ at the point $x \in M$.

In a bit more detail: The proof of Darboux's theorem starts by taking any such f to be our p_1, and then constructs the canonically conjugate generalized coordinate q^1, i.e. the coordinate such that $\{q^1, p_1\} = 1$: so that p_1 generates translation in

the direction of increasing q^1. Indeed the construction is geometrically clear. The symplectic structure means that any such f defines a Hamiltonian vector field X_f, and a flow ϕ^f. We choose a $(2n-1)$-dimensional local submanifold N passing through the given point x, and transverse to all the integral curves of X_f in a neighbourhood of x; and we set the parameter λ of the flow ϕ^f to be zero at all points $y \in N$. Then for any z in a suitably small neighbourhood of the given point x, we define the function $q^1(z)$ to be the parameter-value at z of the integral curve of X_f that passes through z. So by construction, (i) f generates translation in the direction of increasing q^1, and (ii) defining $p_1 := f$, we have $\{q^1, p_1\} = 1$.

This is just the beginning of the proof. But I will not need details of how it goes on to establish the local existence of canonical coordinates, i.e. coordinates such that analogues of (i) and (ii), also for $i \neq 1$, hold. In short, the strategy is to use induction on the dimension of the manifold; for details, cf. e.g. Arnold (1989: 230–232).

To see the significance of this for reducing problems, suppose that there is a constant of the motion, and that we take it as our f, i.e. as the first momentum coordinate p_1. So the system evolves on a $(2n-1)$-dimensional manifold given by an equation $f =$ constant. So writing H in the canonical coordinate system secured by Darboux's theorem, we conclude that $0 = \dot{f} \equiv -\frac{\partial H}{\partial q^1}$. That is, q^1 is cyclic. So as discussed in Section 4.2, we need only solve the problem in the $2n-2$ variables $q^2, \ldots, q^n; p_2, \ldots, p_n$. Having done so, we can find q^1 as a function of time, by solving eq. 4.9 by quadrature.

To put the point in geometric terms:—

(i) The system is confined to a $(2n-1)$-dimensional manifold $p_1 = \alpha =$ constant, M_α say.

(ii) M_α is foliated by a local one-parameter family of $(2n-2)$-dimensional manifolds labelled by values of $q^1 \in I \subset \mathbb{R}$, $M_\alpha = \cup_{q^1 \in I} M_{\alpha, q^1}$.

(iii) Of course, the dynamical vector field is transverse to the leaves of this foliation; i.e. q^1 is not a constant of the motion, $\dot{q}^1 \neq 0$. But since q^1 is ignorable, $\frac{\partial H}{\partial q^1} = 0$, the problem to be solved is "the same" at points x_1, x_2 that differ only in their values of q^1.

6.7 Geometric formulation of the Legendre transformation

Let us round off our development of both Lagrangian and Hamiltonian mechanics, by formulating the Legendre transformation as a map from the tangent bundle TQ to the cotangent bundle T^*Q. In this formulation, the Legendre transformation is often called the *fibre derivative*.

Again, there is a rich theory to be had here. In part, it relates to the topics mentioned in Section 4.2.3: (i) the description of a function (in the simplest case $f : \mathbb{R} \to \mathbb{R}$) by its gradients and axis-intercepts, rather than by its arguments and values; (ii) variational principles. But I shall not go into details about this theory: since this paper emphasises the Hamiltonian framework, a mere glimpse of this theory must suffice. (References, additional to those in Section 4.2.3, include: Abraham and Marsden (1978: Sections 3.6–3.8) and Marsden and Ratiu (1999: Sections 7.2–7.5, 8.1–8.3).)

Let us return to the Lagrangian framework. We stressed in Section 2.2 that a scalar on the tangent bundle, the Lagrangian $L : TQ \to \mathbb{R}$, "determines everything": the dynamical vector field $D =: D_L$; and so for given initial q and $\dot{q}$, L determines a solution, a trajectory in TQ, i.e. $2n$ functions of time $q(t), \dot{q}(t)$ with the first n functions determining the latter.

For the Legendre transformation, the fundamental points are that:

(1) L also determines at any point $q \in Q$, a preferred map FL_q from the tangent space T_q to its dual space T_q^*. Besides this preferred map:

(2) extends trivially to a preferred map from all of TQ to T^*Q; this is the Legendre transformation, understood geometrically;

(3) extends, under some technical conditions (about certain kinds of uniqueness, invertibility and smoothness), so as to carry geometric objects of various sorts defined on TQ to corresponding objects defined on T^*Q, and vice versa.

So under these conditions, the Legendre transformation (together with its inverse) transfers the entire description of the system's motion between the Lagrangian and Hamiltonian frameworks.

I will explain (1) and (2), but just gesture at (3).

(1) Intuitively, the preferred map FL_q from each tangent space T_q to its dual space T_q^* is the transition $\dot{q} \mapsto p$. More precisely: since L is a scalar on TQ, any choice of local coordinates q on a patch of Q, together with the induced local coordinates $q, \dot{q}$ on a patch of TQ, defines the partial derivatives $\frac{\partial L}{\partial \dot{q}}$. At any point q in the domain of the local coordinates, this defines a preferred map FL_q from the tangent space T_q to the dual space T_q^*: $FL_q : T_q \to T_q^*$. Namely, a vector $\tau \in T_q$ with components $\dot{q}^i$ in the coordinate system q^i on Q, i.e. $\tau = \dot{q}^i \frac{\partial}{\partial q^i}$ (think of a motion through configuration q with generalized velocity τ) is mapped to the 1-form whose components in the dual basis dq^i are $\frac{\partial L}{\partial \dot{q}^i}$. That is

$$FL_q : \tau = \dot{q}^i \frac{\partial}{\partial q^i} \ \in \ T_q \quad \mapsto \quad \frac{\partial L}{\partial \dot{q}^i} dq^i \ \in \ T_q^*. \tag{6.33}$$

One easily checks that because the canonical momenta are a 1-form, this definition is, despite appearances, coordinate-independent.

(2) An equivalent definition, manifestly coordinate-independent and given for all $q \in Q$, is as follows. Given $L : TQ \to \mathbb{R}$, define $FL : TQ \to T^*Q$, the *fibre derivative*, by

$$\forall q \in Q, \ \forall \sigma, \ \tau \in T_q : \ \ FL(\sigma) \cdot \tau = \frac{d}{ds} \mid_{s=0} \ L(\sigma + s\tau) \tag{6.34}$$

(We here take σ, τ to encode the identity of the base-point q, so that we make notation simpler, writing $FL(\sigma)$ rather than $FL((q, \sigma))$ etc.) That is: $FL(\sigma) \cdot \tau$ is the derivative of L at σ, along the fibre T_q of the fibre bundle TQ, in the direction τ. So FL is fibre-preserving: i.e. it maps the fibre T_q of TQ to the fibre T_q^* of T^*Q. In local coordinates

$q, \dot{q}$ on TQ, FL is given by:

$$FL(q^i, \dot{q}^i) = \left(q^i, \frac{\partial L}{\partial \dot{q}^i}\right); \text{i.e. } p_i = \frac{\partial L}{\partial \dot{q}^i}. \tag{6.35}$$

An important special case involves a free system (i.e. no potential term in the Lagrangian) and a configuration manifold Q with a metric $g = g_{ij}$ defined by the kinetic energy. (Cf. footnote 2 for the definition of this metric: in short, the constraints being scleronomous (i.e. time-independent, cf. Section 2.1), implies that for any coordinate system on Q, the kinetic energy is a homogeneous quadratic form in the generalized velocities.) The Lagrangian is then just the kinetic energy of the metric,

$$L(q, \dot{q}) \equiv L(\dot{q}) := \frac{1}{2} g_{ij} \dot{q}^i \dot{q}^j \tag{6.36}$$

so that the fibre derivative is given by

$$FL(\sigma) \cdot \tau = g(\sigma, \tau) = g_{ij} \sigma^i \tau^j, \text{ i.e. } p_i = g_{ij} \dot{q}^j. \tag{6.37}$$

(3) We can use FL to pull-back to TQ the canonical 1-form $\theta \equiv \theta_H$ and symplectic form ω from T^*Q (eq. 6.17 and 6.18 with $\omega = -\mathbf{d}\theta$, from Section 6.3.B). That is, we can define

$$\theta_L := (FL)^* \theta_H \text{ and } \omega_L := (FL)^* \omega. \tag{6.38}$$

Since exterior differentiation $\mathbf{d}$ commutes with pull-backs, $\omega_L = -\mathbf{d}\theta_L$. Furthermore:

(i) As one would hope, θ_L, so defined, is Lagrangian mechanics' canonical 1-form, which we already defined in eq. 2.13 (and which played a central role in the Lagrangian version of Noether's theorem).

(ii) One can show that ω_L is non-degenerate iff the Hessian condition eq. 2.3 holds. So under this condition, we can analyse Lagrangian mechanics in terms of symplectic structure.

Given L, we define its energy function $E : TQ \to \mathbb{R}$ by

$$\forall\, v \equiv (q, \tau) \in TQ, \quad E(v) := FL(v) \cdot v - L(v); \tag{6.39}$$

or in coordinates

$$E(q^i, \dot{q}^i) := \frac{\partial L}{\partial \dot{q}^i} \dot{q}^i - L(q^i, \dot{q}^i) \tag{6.40}$$

If FL is a diffeomorphism, we find that $E \circ (FL)^{-1}$ is, as one would hope, the Hamiltonian function $H : T^*Q \to \mathbb{R}$ which we already defined in eq. 4.4.

And accordingly, if FL is a diffeomorphism, then the derivative of FL carries the dynamical vector field $\frac{d}{dt}$ in the Lagrangian description, as defined in eq. 2.8 (Section 2.2, (2)), viz.

$$D_L := \dot{q}^i \frac{\partial}{\partial q^i} + \ddot{q}^i \frac{\partial}{\partial \dot{q}^i}, \tag{6.41}$$

to the Hamiltonian dynamical vector field, viz.

$$D_H := \dot{q}^i \frac{\partial}{\partial q^i} + \dot{p}_i \frac{\partial}{\partial p_i}. \tag{6.42}$$

More generally, one can show if FL is a diffeomorphism, there is a bijective correspondence between the various geometric structures used in the Lagrangian and Hamiltonian descriptions. For precise statements of this idea, cf. e.g. Abraham and Marsden (1978: Theorem 3.6.9) and Marsden and Ratiu (1999: Theorem 7.4.3.), and their preceding discussions.

6.8 Glimpsing the more general framework of Poisson manifolds

Recall that Section 5.1 listed several properties of the Poisson bracket, as defined by eq. 5.3 or 5.6. We end by briefly describing how the postulation of a bracket that acts on the scalar functions $F : M \to \mathbb{R}$ defined on *any* manifold M, and possesses four of Section 5.1's listed properties, provides a sufficient framework for mechanics in Hamiltonian style. The bracket is again called a 'Poisson bracket', and the manifold M equipped with such a bracket is called a *Poisson manifold*.

Namely, we require the following four properties. The Poisson bracket is to be bilinear; antisymmetric; and to obey the Jacobi identity (eq. 5.11) for any real functions F, G, H on M, i.e.

$$\{\{F, H\}, G\} + \{\{G, F\}, H\} + \{\{H, G\}, F\} = 0; \tag{6.43}$$

and to obey Leibniz' rule for products (eq. 5.9), i.e.

$$\{F, H \cdot G\} = \{F, H\} \cdot G + H \cdot \{F, G\}. \tag{6.44}$$

This generalizes Hamiltonian mechanics: in particular, a Poisson manifold need not be a symplectic manifold. The main idea of the extra generality is that the antisymmetric bilinear map that gives the geometry of the state space (the analogue of Section 4.3's symplectic form ω) can be degenerate. So this map can "have extra zeroes", as in eq. 4.37 and 4.38. (This map is induced by the generalized Poisson bracket, via an analogue of eq. 5.7.) This means that a Poisson manifold can have *odd* dimension; while we saw in Section 4.3.3 that any symplectic vector space is even-dimensional—and so, therefore, is any symplectic manifold (Section 6.3.1 and 6.6).

On the other hand, the generalized framework has strong connections with the usual one.[18] One main connection is the result that any Poisson manifold M is a

disjoint union of even-dimensional manifolds, on each of which M's degenerate antisymmetric bilinear form (induced by the generalized Poisson bracket) restricts to be non-degenerate; so that there is an orthodox Hamiltonian mechanics on each such 'symplectic leaf'. Another main connection is that Section 5.3's "one-liner" version of Noether's theorem, eq. 5.18, underpins versions of Noether's theorem for the more general framework.

This generalized framework is important for various reasons; I will just mention two.

(i) For a system whose orthodox Hamiltonian mechanics on a symplectic manifold (dimension $2n$, say) depends on s real parameters, it is sometimes natural to consider the corresponding $(2n + s)$-dimensional space. This is often a Poisson manifold; viz., one foliated into an s-dimensional family of $2n$-dimensional symplectic manifolds. This scenario occurs even for some very familiar systems, such as the pivoted rigid body described by Euler's equations.

(ii) Poisson manifolds often arise in the theory of symplectic reduction. For when you quotient a symplectic manifold by the action of a group (e.g. a group of symmetries of a Hamiltonian system in the sense of Section 6.5), you often get a Poisson manifold, rather than a symplectic one. Indeed, the pivoted rigid body is itself an example of this.

But this generalized framework is a large topic, which we cannot go into: as mentioned, Butterfield (2006) is a philosopher's introduction.

For now, we end with a historical point.[19] It is humbling, but also I hope inspiring, reflection about one of classical mechanics' monumental figures. Namely: a considerable part of the modern theory of Poisson manifolds, including their uses for the rigid body and for symplectic reduction, was already contained in Lie (1890)!

ACKNOWLEDGEMENTS

I am grateful to the editors, not least for their patience; to audiences in Irvine, Oxford, Princeton and Santa Barbara; and to Katherine Brading, Harvey Brown, Hans Halvorson, David Malament, Wayne Myrvold, David Wallace, and especially Graeme Segal, for conversations, comments—and corrections!

NOTES

[1] It is worth noting the point, though I shall not exploit it,that symplectic structure can be seen in the classical solution space of the Lagrangian framework; cf. (3) of Section 6.7.

[2] Though I shall not develop any details, there is of course a rich theory about these and related assumptions. One example, chosen with an eye to our later use of geometry, is that assuming scleronomous constraints, K is readily shown to be a homogeneous quadratic form in the generalized velocities, i.e. of the form $K = \Sigma_{i,j}^{n} a_{ij}\dot{q}^i\dot{q}^j$; and so K defines a metric on the configuration space.

[3] This is not to say that Hamiltonian mechanics makes all problems "explicitly soluble": if only! For a philosophical discussion of the various meanings of 'explicit solution', cf. Butterfield (2004a: Section 2.1).

[4] A note for *afficionados*. Of the three main pillars of elementary differential geometry—the implicit function theorem, the local existence and uniqueness of solutions of ordinary differential equations, and Frobenius' theorem—this paper will use the first only implicitly (!), and the second explicitly in Sections 3 and 4. The third will not be used.

[5] Cf. Brading and Castellani (2003). Apart from papers specifically about Noether's theorem, this anthology's papers by Wallace, Belot and Earman (all 2003) are closest to this paper's concerns.

[6] Here again, 'versions of it' needs scare-quotes. For in what follows, I shall be more limited than these proofs, in two ways. (1): I limit myself, as I did in Section 2.2.1, both to time-independent Lagrangians and to time-independent transformations: so my discussion does not encompass boosts. (2): I will take a symmetry of L to require that L be the *very same*; whereas some treatments allow the addition to L of the time-derivative of a function $G(q)$ of the coordinates q—since such a time-derivative makes no difference to the Lagrange equations.

[7] Other expositions of Noether's theorem for finite-dimensional Lagrangian mechanics include: Arnold (1989: 88–89), Desloge (1982: 581–586), Lanczos (1986: 401–405: emphasizing the variational perspective) and Johns (2005: Chapter 13). Butterfield (2004a, Section 4.7) is a more detailed version of this Section. Beware: though many textbooks of Hamiltonian mechanics cover the Hamiltonian version of Noether's theorem (which, as we will see, is stronger), they often do not label it as such; and if they do label it, they often do not relate it clearly to the Lagrangian version.

[8] An excellent account of this modern integration theory, covering both ordinary and partial differential equations, is given by Olver (2000). He also covers the Lagrangian case (Chapter 5 onwards), and gives many historical details especially about Lie's pioneering contributions.

[9] I have discussed this in terms of some system $(q, \dot{q})$ of coordinates. But the definitions of extensions and displacements are in fact coordinate-independent. Besides, one can show that the operations of displacing a curve within Q, and extending it to TQ, commute to first order in ϵ: the result is the same for either order of the operations.

[10] Since the Lagrangian L is especially associated with variational principles, while the dynamics is given by equations of motion, calling Section 3.2.2's notion 'variational symmetry', and this notion 'dynamical symmetry' is a good and widespread usage. But beware: it is not universal.

[11] All the material to the end of this Subsection is drawn from Brown and Holland (2004a); cf. also their (2004). The present use of the harmonic oscillator example also occurs in Morandi et al. (1990: 203–204).

[12] In the light of this, you might ask about a more restricted implication: viz. must every dynamical symmetry of a set of equations of motion be a variational symmetry of *some or other Lagrangian* that yields the given equations as the Euler-Lagrange equations of Hamilton's Principle? Again, the answer is No for the simple reason that there are many (sets of) equations of motion that are not Euler-Lagrange equations of *any* Lagrangian, and yet have dynamical symmetries.

Wigner (1954) gives an example. The general question of under what conditions is a set of ordinary differential equations the Euler-Lagrange equations of some Hamilton's Principle is the *inverse problem* of Lagrangian mechanics. It is a large subject with a long history; cf. e.g. Santilli (1979), Lopuszanski (1999).

[13] Of course, some aspects of Hamiltonian mechanics illustrate both (i) and (ii). For example, Liouville's theorem on the preservation of phase space volume illustrates both (i)'s integral invariants approach to canonical transformations and (ii)'s connection to statistical mechanics.

[14] But forms are essential for understanding integration over surfaces of dimension two or more: which one needs for the integral invariants approach to Hamiltonian mechanics, and its deep connection with Stokes' theorem.

[15] Details about point transformations on Q defining a canonical transformation on T^*Q, and lifting the vector field X to Γ, can be found: (i) using traditional terms, in Goldstein et al. (2002: 375–376) and Lanczos (1986: Chapter VII.2); (ii) using modern geometric terms (as developed in Section 6), in Abraham and Marsden (1978: Sections 3.2.10–3.2.12) and Marsden and Ratiu (1999: Sections 6.3–6.4).

[16] As announced in Section 2.2.1, I assume the notion of the Lie-derivative, in particular the Lie-derivative of a 2-form. Suffice it to say, as a sketch, that the flow of X defines a map on M which induces a map on curves, and so on vectors, and so on co-vectors, and so on 2-forms such as ω. Nor will I go into details about the equivalence between this definition of X's being symplectic, and X's generating (active) canonical transformations, or preserving the Poisson bracket. For as I have emphasised, I will not need to develop the theory of canonical transformations.

[17] By the way, this Hamiltonian is *not* invariant under boosts. But as I said in Section 2.2.1 and footnote 8, I restrict myself to time-independent transformations; the treatment of symmetries that "represent the relativity of motion" needs separate discussion.

[18] Because of these connections, it is natural to still call the more general framework 'Hamiltonian' as is usually done. But of course this is just a verbal matter.

[19] As mentioned in footnote 10, Olver (2000) gives many details especially about Lie. Cf. in particular Olver (2000: 374–379, 427–428); cf. also Marsden and Ratiu (1999: 336–338, 430–432), and for a full history, Hawkins (2000).

REFERENCES

R. Abraham and J. Marsden (1978), *Foundations of Mechanics*, second edition: Addison-Wesley.

V. Arnold (1973), *Ordinary Differential Equations*, MIT Press.

V. Arnold (1989), *Mathematical Methods of Classical Mechanics*, Springer, (second edition).

G. Belot (2003), 'Notes on symmetries', in K. Brading and E. Castellani (eds.), (2003), *Symmetry in Physics*, Cambridge University Press, pp. 393–412.

K. Brading and E. Castellani (eds.) (2003), *Symmetry in Physics*, Cambridge University Press.

H. Brown and P. Holland (2004), 'Simple applications of Noether's first theorem in quantum mechanics and electromagnetism, *American Journal of Physics* **72**, 34–39. Available at: http://arxiv.org/abs/quant-ph/0302062 and http://philsci-archive.pitt.edu/archive/00000995/.

H. Brown and P. Holland (2004a), 'Dynamical vs. variational symmetries: Understanding Noether's first theorem', *Molecular Physics*, **102**, (11-12 Special Issue), pp. 1133-1139.

J. Butterfield (2004), 'Some Aspects of Modality in Analytical mechanics', in *Formal Teleology and Causality*, ed. M. Stöltzner, P. Weingartner, Paderborn: Mentis.
Available at Los Alamos arXive: http://arxiv.org/abs/physics/0210081 or http://xxx.soton.ac.uk/abs/physics/0210081; and at Pittsburgh archive: http://philsci-archive.pitt.edu/archive/00001192.

J. Butterfield (2004a), 'Between Laws and Models: Some Philosophical Morals of Lagrangian Mechanics'; available at Los Alamos arXive: http://arxiv.org/abs/physics/0409030 or http://xxx.soton.ac.uk/abs/physics/0409030; and at Pittsburgh archive: http://philsci-archive.pitt.edu/archive/00001937/.

J. Butterfield (2004b), 'On Hamilton-Jacobi Theory as a Classical Root of Theory', in A. Elitzur, S. Dolev and N. Kolenda (eds.), *Quo Vadis Quantum Mechanics?*, Springer, pp. 239–273; available at Los Alamos arXive: http://arxiv.org/abs/quant-ph/0210140; or at Pittsburgh archive: http://philsci-archive.pitt.edu/archive/00001193/.

J. Butterfield (2005), 'Between Laws and Models: Some Philosophical Morals of Hamiltonian Mechanics', in preparation.

J. Butterfield (2006), 'On Symplectic Reduction in Classical Mechanics', forthcoming in *The North Holland Handbook of Philosophy of Physics*, ed. J. Earman and J. Butterfield, North Holland.

R. Courant and D. Hilbert (1953), *Methods of Mathematical Physics*, volume I, Wiley-Interscience (Wiley Classics 1989).

R. Courant and D. Hilbert (1962), *Methods of Mathematical Physics*, volume II, Wiley-Interscience (Wiley Classics 1989).

E. Desloge (1982), *Classical Mechanics*, New York: John Wiley.

J. Earman (2003), 'Tracking down gauge: an ode to the constrained Hamiltonian formalism', in K. Brading and E. Castellani (eds.) (2003), *Symmetry in Physics*, Cambridge University Press, pp. 140–162.

H. Goldstein et al. (2002), *Classical Mechanics*, New York, Addison-Wesley (third edition).

T. Hawkins (2000), *Emergence of the Theory of Lie Groups: An Essay in the History of Mathematics 1869–1926*, New York: Springer.

M. Henneaux and C. Teitelboim (1992), *Quantization of Gauge Systems*, Princeton University Press.

O. Johns (2005), *Analytical Mechanics for Relativity and Quantum Mechanics*, Oxford University Press.

J. José and E. Saletan (1998), *Classical Dynamics: a Contemporary Approach*, Cambridge University Press.

H. Kastrup (1987), 'The contributions of Emmy Noether, Felix Klein and Sophus Lie to the modern concept of symmetries in physical systems', in *Symmetries in Physics (1600–1980)*, Barcelona: Bellaterra, Universitat Autonoma de Barcelona, pp. 113–163.

S. Lie (1890). *Theorie der Transformationsgruppen: zweiter abschnitt*, Leipzig: B.G.Teubner.

C. Lanczos (1986), *The Variational Principles of Mechanics*, Dover; (reprint of the 4th edition of 1970).

J. Lopuszanski (1999), *The Inverse Variational Problem in Classical Mechanics*, Singapore: World Scientific.

J. Marsden and T. Ratiu (1999), *Introduction to Mechanics and Symmetry*, second edition: Springer-Verlag.

G. Morandi et al. (1990), 'The inverse problem of the calculus of variations and the geometry of the tangent bundle', *Physics Reports* **188**, 147–284.

P. Olver (2000), *Applications of Lie Groups to Differential Equations*, second edition: Springer-Verlag.

R. Santilli (1979), *Foundations of Theoretical Mechanics, vol. I*, New York: Springer-Verlag.

D. Wallace (2003), 'Time-dependent Symmetries: the link between gauge symmetries and indeterminism', in K. Brading and E. Castellani (eds.), (2003), *Symmetry in Physics*, Cambridge University Press, 163–173.

E. Wigner (1954), 'Conservation laws in classical and quantum physics', *Progress of Theoretical Physics* **11**, 437–440.

WILLIAM DEMOPOULOS*

4. ON THE NOTION OF A PHYSICAL THEORY OF AN INCOMPLETELY KNOWABLE DOMAIN

For Jeffrey Bub on the occasion of his 63rd birthday

ABSTRACT

How might a physical theory have the consequence that facts about some of the things it seeks to describe cannot, as a matter of principle, be completely known? The paper articulates the components of the conceptual structure of a theory that is capable of exhibiting such "inherent incompleteness." Although the framework of the discussion is indebted to axiomatic quantum logic, the analysis is at variance with the quantum logical interpretation of quantum mechanics.

1 INTRODUCTION

How might we conceptualize a physical theory, one of whose principal consequences is that our knowledge of the objects with which it deals is necessarily incomplete? What is the nature of such incompleteness, and would it allow for a sense in which the theory of an incompletely knowable domain is itself complete? To address these questions, it is necessary to explore various components of the conceptual structure of a theory that might exhibit such "inherent incompleteness." In particular, we will want to know what, for such a theory, constitutes a representation of the facts that are incompletely known, and what constitutes a representation of the knowledge of them that is theoretically possible; we will require an account of the sense in which this knowledge fails to be complete, and an explanation of the basis for the failure of completeness.

My approach to these issues would hardly have suggested itself without the foundational investigations of the 1960s, especially those of Kochen and Specker (1967) and their reconsideration of von Neumann's (1932) proof that quantum mechanics cannot be supplemented with hidden variables. Although the framework of the present paper is taken from these highly specific foundational studies, the issues with which it deals should be of broader interest—especially if it can be shown that there are actual examples of fundamental theories that are empirically successful but inherently

* Dept. of Logic and Philosophy of Science, University of California/Irvine, Irvine CA 92697–5100 USA and Dept. of Philosophy, University of Western Ontario, London N6A 3K7, Canada; E-mail: wgdemo@uwo.ca.

W. Demopoulos and I. Pitowsky (eds.), Physical Theory and its Interpretation, 101–116.

incomplete. The notion that a fundamental science like physics cannot take for granted our epistemic relation to the world and that it must therefore define it—articulate a framework within which it is coherently expressed—underlies the preeminent position of the analysis of space and time in Newtonian mechanics and relativity.[1] The epistemological analysis of the general character of measurement plays a comparable role in quantum mechanics and the analysis of inherent incompleteness.

The following paper extends my (2004) in several respects. Here, as in the earlier paper, elementary propositions are central to the analysis and are taken to comprise a physical theory's basic representational apparatus; all its other representational devices are, in a way I will soon explain, derivative from them. The present paper's distinction between basic and derived structures of propositions replaces an earlier and less precise terminology, and this has made possible a sharper formulation of inherent incompleteness. The connection with the quantum mechanical description of spin has also been explicitly drawn.

I owe Jeffrey Bub an enormous debt, first for introducing me to the subject almost 40 years ago, and secondly, for sustaining my interest in it over the course of a long and rewarding friendship. The discussion which follows has been influenced throughout by the recent work of my friend and co-editor Itamar Pitowsky.

2 THE REPRESENTATION OF ELEMENTARITY

I will be almost exclusively concerned with a special class of physical propositions, namely propositions that ascribe direction-dependent properties to physical systems; in the simplest case, such propositions form a class $\boldsymbol{P}$ of propositions P_x, for x a direction (or ray) through a point of ordinary physical space. It is clear that, except for the choice of rays in E^3 as an index set, this notion of proposition is a highly abstract and formal one. It is, nevertheless, suggestive of the applications I will be considering, and it is sufficiently complex to allow the formulation of issues that will be among our main concerns.

In order to speak of the elements of $\boldsymbol{P}$ as propositions, there are certain minimal conditions which the system they comprise—the "logical space" in which they lie—must satisfy: First, complements must be defined, and this means that there must be maximum and minimum propositions—0 and 1, respectively—such that $P_x \vee \neg P_x = 1$ and $P_x \wedge \neg P_x = 0$. Secondly, for the P_x to belong to a common space of propositions, 0 and 1 must be unique, so that $P_x \vee \neg P_x = P_y \vee \neg P_y = \ldots 1$ and $P_x \wedge \neg P_x = P_y \wedge \neg P_y = \ldots 0$. We will assume that this is the case. These appear to be the minimal logical-combinatorial assumptions—existence of complements, maximum and minimum elements—one might impose on $\boldsymbol{P}$ if it is to form a class of objects that may plausibly be regarded as propositions.

Since the elements of $\boldsymbol{P}$ are not only propositions, but propositions of an empirical theory, we will assume that there is an ideal operational or measurement procedure (these terms are used interchangeably) associated with each P_x in $\boldsymbol{P}$. Such a procedure is ideal in the sense that it constitutes a theoretically, if not practically, feasible criterion of application for the property contained in P_x. The measurement

procedure is criterial in the sense that, should its application fail to find that P_x holds, we may conclude that $\neg P_x$ holds. Thus, excluded middle holds for such propositions and measurement procedures, in contrast with the case of constructive proofs and mathematical propositions where it may happen that there is neither a constructive proof of p nor a constructive proof that every proof of p can be transformed into a proof of the falsum $\perp$—which is why excluded middle is rejected as a principle of intuitionist logic.

Our focus is on an abstract aspect of measurement, namely on the simultaneous measurability of directional properties that are constituents of elementary physical propositions. What can we say of the simultaneous measurability of such properties—and derivatively, of the propositions in which they occur—without prejudging empirical issues regarding the character of this relation?

Since every measurement must "find" 1 and "exclude" 0, the propositions, P_x, $\neg P_x$, 0, 1 must be simultaneously measurable, or, as I will say, *comeasurable*. This much seems to follow from our understanding of the comeasurability of elementary propositions with an empirical content, however comeasurability is spelled out in physical detail. Whenever a pair of propositions is comeasurable, lattice operations of meet and join are defined for them, since, if the same ideal measurement decides both propositions, it also decides their conjunction (meet) and disjunction (join).

In a general or abstract consideration of propositions, the existence of meets and joins is always naturally assumed as a matter of course. We however are considering a special class of propositions, those for which it is intuitively natural to inquire about their comeasurability and, perhaps, to discover that there is no theoretically specifiable operational procedure which simultaneously decides their constituent properties. Recall that a pair of propositions is comeasurable when there is a *single* theoretically specifiable ideal measurement procedure which is criterial for the properties that are constituent in both of them. Since the conjunction and disjunction of noncomeasurable propositions cannot be associated with a single ideal measurement procedure, they are excluded from our study, and the lattice operations of meet and join are treated as partial operations, defined *only* for comeasurable pairs of propositions.

From a classical or Boolean algebraic point of view, the totalness of the lattice operations is taken for granted. This is perhaps because the notion of proposition from which the classical perspective begins is an abstract or general one which, unlike the special case of elementary physical propositions, is not necessarily tied to the notion of an operational procedure. But once the association of elementary propositions with ideal measurement procedures has been made, and comeasurability explicitly recognized, it is evidently possible that the *only* sets of comeasurable propositions are those of the form $\{P_x, \neg P_x, 0, 1\}$ or $\{0, 1\}$—in which case lattice operations of meet and join need not be defined on subsets of $\boldsymbol{P}$ larger than $\{P_x, \neg P_x, 0, 1\}$. When the existence of meets and joins is inferred from the assumption that they are defined for comeasurable propositions, the Boolean framework presents itself as one that is distinguished by the fact that it is *maximally* committal regarding the extension of the comeasurability relation, taking it in fact to be the universal relation. This is certainly one way of proceeding, but it is by no means the only one. And indeed, there is a

minimally committal alternative, namely one that assumes not a Boolean algebra, but the partial Boolean algebra whose family of Boolean subalgebras consists of just the two element Boolean algebra and algebras of the form $\{P_x, \neg P_x, 0, 1\}$. A priori, these are the two extreme cases that present themselves: either the comeasurability relation is the universal relation on elementary propositions or it is "smallest possible." Without additional considerations pertaining to the specific character of the propositions under investigation, it is difficult to see how one might motivate a choice of domain of definition for meet and join that lies between the largest and smallest comeasurability relations. Later we will see how empirical considerations may be brought to bear on the structure of comeasurability.

When the comeasurability relation is assumed to have the smallest possible extension, the logical structure of the family $\boldsymbol{P}$ of elementary propositions is extremely simple and is represented by the free partial Boolean algebra $\mathrm{B}(\mathrm{E}^2)$—the so-called "2-dimensional case" comprised of subspaces of the Euclidean plane, with $x \wedge y$ the intersection of subspaces and $x \vee y$ their span, $x^{\perp}$ the orthogonal complement of x, 0 and 1 (respectively) the empty subspace and the whole plane. The maximal Boolean subalgebras of $\boldsymbol{P}$ are composed of elements, P_x, their complements, $\neg P_x$, and 0 and 1. The algebra is freely generated by the P_x in the sense that any map from them to a partial Boolean algebra can be extended to a homomorphism. More precisely, for each x in E^3, let P_x^* be either P_x or $\neg P_x$. Then $\{P_x^* : x \text{ in } \mathrm{E}^3\}$ is an *independent* set of elements in the sense that every map from $\{P_x^* : x \text{ in } \mathrm{E}^3\}$ to a partial Boolean algebra can be extended to a homomorphism. Notice that the property by which we have defined the notion of an independent set of elements is usually derived as a theorem from a definition expressed in terms of meets and complements; this shows that although such a definition is not available to us because of the partialness of the algebra, the concept of independence remains a meaningful one in the present context.

3 THE TRACTARIAN NOTION OF ELEMENTARITY

It is a curious consequence of our analysis that the elementary propositions of $\boldsymbol{P}$ fulfill *all* the requirements that, in the *Tractatus*, Wittgenstein appears to have demanded of the notion of an elementary proposition. As we will see, Wittgenstein's requirements are unsatisfiable in all but the simplest classical logical examples, a situation that has prevented even the appearance of there being an interesting application of the Tractarian notion of elementarity.

In developing his account of an elementary proposition, Wittgenstein's goal was to give a completely "combinatorial" analysis of the notions of logical possibility and necessity. On such an analysis, the compatibility and incompatibility of propositions would be discoverable on the basis of their constituent logical forms. Now algebraic atoms are minimal non-zero elements. In a Boolean algebra, the conjunction of two atoms is always defined and is the zero of the algebra; hence any two atoms are logical contraries of one another. It follows that in a Boolean algebra elementary propositions cannot, in general, be algebraic atoms and form an independent set of

generators of the algebra. The exception is the four element Boolean algebra, where $\{p\}$ and $\{\neg p\}$ are independent sets of generators, both p and $\neg p$ are atoms, and p and $\neg p$ can be held to exclude one another on the basis of their logical form. In every larger Boolean algebra there will exist atoms which, though they are logical contraries, are not logical complements. It is clear, in this case, why p excludes $\neg p$, but it is evident that in general if p and q are elementary they cannot exclude one another since there is nothing about them to which *logic* might appeal in order to explain the fact that they cannot be true together—if indeed they cannot be true together. But it seems equally clear that if there are to be actual examples of elementary propositions, they must sometimes be logical contraries of one another. This was the impasse that ultimately led Wittgenstein in his (1929) to abandon the Tractarian theory of elementary propositions.

Considered in isolation, the requirement that elementary propositions should be algebraic atoms has little intrinsic plausibility. It does however fall out as a *consequence* of the present analysis of elementarity that elementary propositions are also algebraic atoms. This is made possible by the presence in the analysis of the relation of comeasurability. The propositions we have described are elementary in the sense that each is directly associated with an ideal operational procedure specifying the criterion of application for the property it contains. Operational procedures bring with them the relation of comeasurability. And since we require of every proposition that it be associated with a measurement procedure that simultaneously decides the properties it contains, only comeasurable propositions have a lattice meet and join. The argument that $\mathrm{B}(\mathrm{E}^2)$ correctly represents the logical structure of the elementary propositions is based on the premise that the comeasurability relation should be the minimal comeasurability relation: since we cannot be held to know a priori how extensive comeasurability is, we should make the weakest assumption about its extension that is compatible with $\boldsymbol{P}$ being a family of propositions. But the propositions of $\mathrm{B}(\mathrm{E}^2)$ are elementary in two respects: They *freely generate* the algebra and thus exhibit elementarity in the sense of comprising independent sets of building blocks out of which the whole algebra is constructed. In addition, they are elementary in the sense that they are algebraic *atoms*, i.e., minimal non-zero elements of the algebra. The partialness of the algebra thus allows for elementary *atomic* generators that are also *independent* sets of generators, a situation that can arise in the classical context only in the case of a four-element Boolean algebra. The possibility of algebras of propositions that contain atoms which do not exclude one another, and are therefore not orthogonal, is arguably the central conceptual innovation that the foundations of quantum mechanics holds for the study of elementary propositions.

4 STATISTICAL STATES AND THE GEOMETRY OF OPERATIONAL PROCEDURES

$\boldsymbol{P}$ is an algebra of propositions isomorphic to $\mathrm{B}(\mathrm{E}^2)$ in which (i) knowledge of one proposition P_x implies literally nothing about knowledge of another proposition $P_y, x \neq y$, and (ii) unlike the Boolean case, states of complete information regarding

all the P_x do not occur in the algebra. The information given by the relatively impoverished logical-combinatorial structure must be supplemented in two important ways: first, we require the inclusion of a class $\boldsymbol{S}$ of statistical states ψ and an algorithm $prob_\psi$ for assigning probabilities to the P_x on their basis, and secondly, we require a more fine-grained analysis of the nature of the possible ideal measurement procedures for the constituent properties of elementary propositions. Let us first consider the effect of introducing a set of statistical states.

Given $\boldsymbol{S}$ and $prob_\psi$ we can, following the approach to axiomatic foundations of Mackey (1963) and others, define a partial order relation $\leq$ on the P_x by putting $P_x \leq P_y$ if for every statistical state ψ, $prob_\psi(P_x) \leq prob_\psi(P_y)$. The complement $\neg P_x$ of P_x is defined by the condition that for all $\psi, prob_\psi(\neg P_x) = 1 - prob_\psi(P_x)$. Once the comeasurability relation has been specified, we can define partial lattice operations of meet and join in terms of this ordering—or, more accurately, in terms of the ordering it induces on equivalence classes of elementary propositions, equivalent with respect to the relation $P_x \approx_{\boldsymbol{S}} P_y$ iff $P_x \leq P_y$ and $P_y \leq P_x$.

A very simple example—essentially, the quantum mechanical description of a spin-1/2 system—illustrates this idea. Given two opposite directions $x = x+$ and $x-$ along a ray x in E^3, together with associated elementary propositions P_x and P_{x-}, suppose that for every statistical state $\psi, prob_\psi(P_x) = 1 - prob_\psi(P_{x-})$, so that $\neg P_x = P_{x-}$. The information which the states in $\boldsymbol{S}$ tell us one proposition yields regarding another is represented by $B(H^2)$—the partial Boolean algebra of subspaces of a complex Hilbert space of two dimensions—which, as a partial Boolean algebra, is isomorphic to $B(E^2)$. In this example it is assumed that the operational procedure associated with a pair P_x and P_{x-} of elementary propositions is one whose geometric characterization depends on a specification of the orientation of the apparatus along a direction in E^3. In this simplest case all that changes when we pass from the purely logical-combinatorial representation given by $B(E^2)$ to one that is informed by the probability algorithm and set of statistical states is the identification of $\neg P_x$ as P_{x-}. The algebraic structure of the representation is unchanged.

The following terminological conventions will be useful: Let us call the partial Boolean algebra $\boldsymbol{P}$ of elementary propositions *the basic structure*; the quotient structure $\boldsymbol{P}/\approx_{\boldsymbol{S}}$, with $\leq, \neg, \wedge$ and $\vee$ defined in terms of $prob_\psi$ and $\boldsymbol{S}$, we will call *a derived structure* of elementary propositions or *the derived structure based on* $\boldsymbol{S}$. By a convenient extension of terminology, elements of the basic structure are *basic elementary propositions*, those of a derived structure, *derived elementary propositions*. The "properly" elementary propositions are the elements of the basic structure; the elements of a derived structure are abstractions from basic elementary propositions to equivalence classes of them.

The basic structure is *always* assumed to be isomorphic to $B(E^2)$. There is nothing to exclude the possibility that the basic and derived structures are algebraically the same, but it can also happen that they are very different, in which case, a consequence of combining $\boldsymbol{P}$ with a family of statistical states is that the character of the comeasurability relation on the associated derived structure changes from the minimal comeasurability relation of $B(E^2)$ to something more complex. This brings us to

our second example and an important refinement in the analysis of ideal operational procedures.

All the propositions we are considering involve the ascription of direction-dependent properties that are accessible to us only through their associated criteria of application—by their ideal operational procedures. The case described earlier was especially simple, involving criteria of application that distinguish between positive and negative directions along a ray. We want now to consider the case of propositions involving properties whose attribution requires ideal measurement procedures that involve decompositions of the rays of E^3 into orthogonal triples.

It is evident that the identity of a ray in physical space does not depend on whether the plane to which it is orthogonal is represented as the span of one or another pair of mutually orthogonal rays. Our conception of the geometry of the rays through a point of space is such that the identity of a ray is independent of the choice of basis to which it belongs. But in the case of the algebra of propositions, it is an open question to what extent the independence of a ray from a basis is inherited by the direction-dependent properties and propositions of ***P***. To address this question we need to consider a more complex example and a correspondingly more complex notation since we must indicate both the ray associated with the constituent property of the proposition and the decomposition that characterizes the property's criterion of application.

We will begin with a purely abstract description. Let (x, y, z) be a decomposition of E^3 into an orthogonal triple of rays. Let $x\theta = x\theta(x, y, z) = (x, \theta y, \theta z)$ be the decomposition that results from a rotation about x through an angle θ sending y to θy and z to θz. Clearly, for $0 \leq \theta < 90$ the decompositions of E^3 are all distinct and contain the ray x, and for $\theta = 0, x\theta = (x, y, z)$. For $\theta \neq \theta'$, the propositions $P_{x\theta}$ and $P_{x\theta'}$ are associated with the same *ray* in E^3, but with measurement procedures that are distinguished by the different decompositions in terms of which they are characterized. The distinctness of the measurement procedures means that $P_{x\theta}$ and $P_{x\theta'}$ may involve distinct properties, making them distinct as propositions. This is in contrast to the case we considered first, where the distinctness of propositions is exhausted by the distinctness of the rays with which they are associated or by the difference of direction along a ray.

In the present example, we have a family of propositions $P_{x\theta}$ whose associated direction-dependent properties have as their criteria of application ideal measurement procedures. Each measurement procedure involves a decomposition of E^3 into an orthogonal basis. Although it may seem artificial to distinguish $P_{x\theta}$ and $P_{x\theta'}$ it is important to bear in mind that we are seeking to isolate the minimal initial assumptions that are required in order for the $P_{x\theta}$ to count as a class of propositions. These assumptions may later be supplemented on the basis of further empirical considerations in a way we have yet to explore, but that will not affect their status as a starting point; they represent the minimal logical and empirical assumptions we require in order to have a family of propositions at all.

Thus, in the case of *basic* propositions, the initial constraints on the relation of comeasurability are determined purely by what the analysis of the notion requires: if comeasurability is understood in terms of ideal measurements that are *criterial* for the

property mentioned in P_x, then the constraints that preanalytic intuition imposes on the relation can be satisfied if it is identified with the smallest comeasurability relation. But what should the relation be based upon in the case of a *derived* structure? For a class of basic propositions associated with a spin-1 particle, it is an empirical fact that there is a distinguished family of ideal operational procedures, each of which is specified by an orthogonal decomposition of E^3, representing the axes of the coordinate frame associated with the measurement apparatus of the operational procedure. It is also an empirical fact that for triples of directional properties in each of three mutually orthogonal directions of E^3 there is a *single* ideal operational procedure which simultaneously measures all three properties. The associated propositions, P_x, P_y, P_z, are therefore comeasurable. The derived structure of interest to us arises from the empirical fact that the probability assignments determined by the ψ in $\boldsymbol{S}$ are independent of the "measurement context," where a difference in measurement context is produced by a rotation in E^3 of the coordinate frame of the measurement apparatus about one of its principal axes. More specifically, let (x, y, z) be an orthogonal triple of rays in E^3 and let $P_{x\theta}, P_{y\theta}, P_{z\theta}, 0 \leq \theta < 90$ be elementary propositions associated with x, y and z, respectively. There are three classes of ideal measurement procedures for the constituent properties: $\{(x, \theta y, \theta z) : 0 \leq \theta < 90\}$, $\{(\theta x, y, \theta z) : 0 \leq \theta < 90\}$, and $\{(\theta x, \theta y, z) : 0 \leq \theta < 90\}$. The statistical states ψ in $\boldsymbol{S}$ are such that for every choice of θ, the propositions $P_{x\theta}$ are $\approx_{\boldsymbol{S}}$-equivalent, as are the propositions $P_{y\theta}$, and the propositions $P_{z\theta}$. This holds independently of the initial choice of (x, y, z). Hence the correspondence from E^3 to $\boldsymbol{P}/\approx_{\boldsymbol{S}}$ which sends each ray x to its derived proposition $[P_{x\theta}]$ is one-one. This justifies dropping the more complex notation and writing P_x for the derived proposition $[P_{x\theta}]$ of $\boldsymbol{P}/\approx_{\boldsymbol{S}}$; i.e., we see that in this example, the *derived* propositions are merely *direction*-dependent.

If ideal operational procedures are parameterized by orthogonal decompositions of E^3 and the probability assignments determined by the states ψ in $\boldsymbol{S}$ exhibit the rotational symmetry we have described, then the statistical states ψ in $\boldsymbol{S}$ are such that for every choice of θ, the propositions $P_{x\theta}$ are $\approx_{\boldsymbol{S}}$-equivalent, as are the propositions $P_{y\theta}$, and the propositions $P_{z\theta}$. In this example, $\{P_x, P_y, P_z\}$ is a comeasurable family of propositions whenever (x, y, z) is an orthogonal triple of rays, and with respect to the ordering $\leq$ given by $prob_\psi$, the partial operations $\wedge$ and $\vee$ are such that $\neg P_x = P_y \vee P_z$, $\neg P_y = P_x \vee P_z$ and $\neg P_z = P_x \vee P_y$, and dually, i.e. after interchange of $\vee$ for $\wedge$ and complemented for uncomplemented propositions. The derived algebra generated by all such triples is isomorphic to the partial Boolean algebra of subspaces of E^3, where the Boolean subalgebras generated by the sets $\{P_x, P_y, P_z\}$ are *maximal* comeasurable subsets of $\boldsymbol{P}/\approx_{\boldsymbol{S}}$. This is the structure that, as a partial Boolean algebra, is shared by the derived structure $\boldsymbol{P}/\approx_{\boldsymbol{S}}$ of our example. This is a much more complex object than $B(E^2)$, and as we will soon see, this complexity is reflected in the representation of how knowledge of one proposition bears on knowledge of another. The character of the derived structure of this example is thus a consequence of the rotational symmetry of the states in $\boldsymbol{S}$. The invariance of probability under rotation of the measurement apparatus about an axis is the central principle on which our analysis is based, a principle that has the epistemic status of a broadly confirmed empirical hypothesis.

Although the quantum mechanical description of a spin-1/2 system is easily read off our first example, locating the spin-1 description in our second example is somewhat less straight-forward. The canonical measurement procedure associated with the property of spin is a Stern-Gerlach apparatus. In the case of a spin-1 system, there are three possible outcomes to a measurement of spin in a particular direction. The outcomes are associated with propositions of the form:

The spin in the direction x is i ($i = 1, 0, -1$).

For $i = 0$, the propositions

Q_x The spin in the direction x is zero

are statistically equivalent to the propositions P_x of our example. This justifies the interpretation of the propositions P_x as

P_x The square of the spin$_x$is zero.

More precisely, the *basic* elementary propositions of the example are of the form

$P_{x\theta}$ The square of the spin$_{x\theta}$ is zero,

where θ specifies the orientation of the measurement apparatus which simultaneously measures $P_{x\theta}, P_{\theta y}$ and $P_{\theta z}$. What one finds is that for all statistical states ψ of the quantum mechanical description of such a system, and for all θ and θ', $prob_\psi(P_{x\theta} = prob_\psi(P_{x\theta'}) = prob_\psi(Q_x)$—a fact which we represent by dropping the reference to a particular decomposition and writing

P_x The square of the spin$_x$ is zero

for the *derived* propositions based on $\boldsymbol{S}$.

The initial or basic structure consists of propositions whose constituent properties have associated with them measurement procedures that are maximal with respect to the number of *square of the spin$_x$* properties they simultaneously decide. The specification of such a procedure necessarily involves a decomposition of E^3 into orthogonal triples of directions. In the case of some properties, the canonical measurement procedure is such that it suffices to mention just the direction itself. This is illustrated by the example of *spin$_x$* in the example of a spin-1/2 system, and it holds as well for *spin$_x$* for a spin-1 system (the Q_x above). But in other cases—and in particular, in the case of interest to us, namely, *square of the spin$_x$* for a spin-1 particle—this is not true, and different decompositions, corresponding to different maximal measurement procedures, must be included in the designation of the propositions. The salient difference among different ideal operational procedures is therefore captured by different choices of θ. The context-independence of probability is expressed by the invariance of the statistical distribution under change of θ. The probabilities—and therefore the derived propositions—exhibit a rotational symmetry that the basic propositions do not share. The differences among operational procedures that we have isolated therefore

have a special character since they depend only on the orthogonal decompositions of E^3 with which they are associated. This is to be contrasted with the more general case, where differences among ideal measurement procedures may not be susceptible of a simple and unitary theoretical—let alone geometrical—characterization.

On the present reconstruction of the framework of the non-relativistic theory, the source of the Hilbert space formalism is located in an invariance principle that concerns the statistical equivalence of the elements of a canonical class of basic elementary propositions, their ideal operational procedures, and the geometry of ordinary space. The classical or macroscopic level enters the analysis and representation of measurement through the rotation group of E^3, which parametrizes measurement procedures for the class of directional properties of interest. The dependence of probability on the geometry of space is the prototype for the general case: the Hilbert space formalism abstracts from the dependence of probability on the *Euclidian* angle between the directions involved in directional properties of the sort considered here to the dependence on the angle in *Hilbert* space that relates the subspaces by which basic elementary propositions—of whatever character—are represented.

5 INHERENT INCOMPLETENESS

Recall that in order for a theory to exhibit the essential or inherent incompleteness we are attempting to elucidate, it must contain (i) a representation of the facts which are incompletely known, (ii) a representation of the knowledge of them that is theoretically possible, (iii) an account of the sense in which this knowledge fails to be complete, and (iv) an explanation of the basis for the failure of completeness. For the special class of elementary propositions we are considering, facts correspond to true elementary propositions drawn from the basic structure. Since our knowledge of the propositions of the basic structure must be compatible with our available statistical information, what can be known of these propositions is constrained by the appropriate derived structure, which is itself determined by a set of statistical states and probability algorithm.

I will say that a derived structure is an *encoding* of the information a set of statistical states contains about the basic elementary propositions of $\boldsymbol{P}$ if the ψ in $\boldsymbol{S}$ together with $prob_\psi$ yield exactly the generalized probability measures definable on $\boldsymbol{P}/\approx_{\boldsymbol{S}}$.[2] A derived structure is the appropriate vehicle for addressing the question of completeness if, and only if, it encodes the available statistical information.

Our knowledge of $\boldsymbol{P}$ is *inherently incomplete* when the statistical information contained in the set of statistical states and probability algorithm are encoded by a derived structure for which there are no generalized 2-valued measures, where a *generalized 2-valued measure* is a generalized probability measure taking values in $\{0, 1\}$.

It is a consequence of a celebrated theorem of Gleason (1957) that the partial Boolean algebras $B(H^n)_{n\geq 3}$ encode the statistical states of quantum mechanics, a fact that is often understood to mean that the theory is "complete" in the sense that its probability algorithm $prob_\psi$ generates all possible generalized probability measures on the appropriate structure. But what is of special importance to us is that it is a

feature of the propositions of our second example that knowledge of them—even of a finite subset of them—is *inherently incomplete*. The force of *inherent* incompleteness emerges as follows: To say that $\boldsymbol{P}/\approx_{\boldsymbol{S}}$ has no 2-valued measure is not merely to claim that there are no such states in $\boldsymbol{S}$; rather, the point is that the states that are in $\boldsymbol{S}$ are such that the structure they force for $\boldsymbol{P}/\approx_{\boldsymbol{S}}$ is *logically incompatible* with the existence of a generalized 2-valued measure on $\boldsymbol{P}/\approx_{\boldsymbol{S}}$.

The nature of the notion of incompleteness we have uncovered is well illustrated by a theorem of Pitowsky (1998): Given any two noncomeasurable propositions P_x and P_y represented by rays in $\mathrm{B}(\mathrm{H}^3)$, we can always find a *finite* set Γ of rays of $\mathrm{B}(\mathrm{H}^3)$ which contains P_x and P_y and has an orthogonality structure that forces any 2-valued measure on Γ to assign them both 0. More generally (see Pitowsky 2005), when the derived structure of propositions has the character of our more elaborate example, either the probability of any two noncomeasurable P_x and P_y is zero, or at least one has a probability strictly between zero and one.[3] Hence, the inherent incompleteness of our knowledge of the propositions in $\boldsymbol{P}$ does not depend on there being a continuum—or even a countably infinite—number of them.

Notice that the sense of incompleteness with which we are concerned is one that is *internal* to a theory: incompleteness is relative to a theory's specification of a family of elementary propositions and its characterization of the available statistical information regarding them. In particular, the internal character of incompleteness means that it does not involve a claim which quantifies over all possible theories of the properties belonging to the propositions of $\boldsymbol{P}$. This stands in marked contrast to Heisenberg's early views (1927) concerning the ineradicable disturbance a measurement of a quantum mechanical system produces.[4] Heisenberg treats such systems as incompletely knowable, but the notion of incompleteness to which his account appeals is not an internal one precisely because it does quantify over all possible theories of the measurement process. In other respects, the present view has a certain affinity with Heisenberg's account, especially if we see it as an attempt to articulate a theory that is minimalist in its nonempirical commitments concerning the scope of the comeasurability relation. There is another connection with Heisenberg that we are not yet in a position to address, but which we will come to shortly.

It is solely a consequence of the empiricism of the framework we are articulating that it distinguishes between $P_{x\theta}$ and $P_{x\theta'}$ when the decompositions of E^3 effected by their associated criteria of application differ. An economy arises when, under pressure of experience, distinctions among propositions are collapsed by placing statistically equivalent propositions into the same equivalence class. For the present analysis, the fact that the result of dividing $\boldsymbol{P}$ by $\approx_{\boldsymbol{S}}$ is typically an object of very different form from $\boldsymbol{P}$ is of far greater interest than the economy effected by the "identification" of statistically equivalent basic propositions. The difference in form has the consequence that statistically equivalent propositions must sometimes differ in truth value, so that our best statistical information can be logically incompatible with knowing the truth values of all elementary propositions. This brings out a fundamental difference in perspective between the present, epistemic, approach and modal interpretations of

quantum mechanics (see e.g. (Bub 1997)). A modal interpretation seeks to characterize certain maximal but *proper* subsets of propositions of $B(H^3)$—and hence, of ***P***—for which there exist 2-valued measures. Such subsets can include noncomeasurable propositions without contradicting Pitowsky's theorem because it is not required that all the propositions in the set Γ of the theorem belong to one of the maximal subsets of a modal interpretation. For a modal interpretation, facts correspond to the true propositions of such a maximal subset, but many propositions are neither true nor false. In modal interpretations, there is therefore nothing corresponding to the incompleteness of our knowledge of what is and what is not the case because there is simply nothing to know.

6 QUANTUM LOGIC AND HIDDEN VARIABLES

Although the free partial Boolean algebra on a continuum of generators has many strikingly non-classical features—it is, for example, *irreducible* in the sense that only 0 and 1 are comeasurable with every element of the algebra—it violates no law of classical logic. The character of its departure from classical ideas is, therefore, one that does not carry with it a new conception of truth. The present approach is able to preserve the determinacy of truth value for elementary propositions because the basic structure in which they lie, though not a Boolean algebra, is embeddable into a Boolean algebra, and therefore has a plethora of 2-valued homomorphisms. It follows from this that all classical tautologies hold in $B(E^2)$ under a suitably generalized sense of propositional validity (cf. Kochen and Specker (1967) Theorem 4). It can therefore be maintained that the notion of truth on which the account relies is the classical one since both it and the classical notion obey the same "laws of truth."

The quantum logical interpretation of quantum mechanics of Bub (1974), Demopoulos (1977) and Friedman and Putnam (1978) is based on the idea that every elementary proposition is determinately true or false in a much broader range of cases than when the basic structure which contains them is $B(E^2)$. Even when the logical structure of elementary propositions is represented by $B(E^3)$, it follows on the quantum logical interpretation that every proposition is determinately true or false, since for every proposition P, $P \vee \neg P$ is always the unit of $B(E^3)$, and is therefore *true*. If this interpretation could be sustained, it would have the advantage of securing determinacy of truth value without the context-dependence of basic propositions: propositions that are merely statistically equivalent on the view we have been developing would actually be the *same* proposition, despite the association of their constituent properties with diverse criteria of application. The difficulty, however, is that it is unclear how to explain the notion of truth the interpretation requires. Consider, for example, Kochen and Specker's proof that there is a finite family of propositions that have no 2-valued measure. The orthogonality graph of the rays used by Kochen and Specker in their proof can be represented by a propositional formula, $\varphi = \varphi(x_1, \ldots, x_{86})$; when interpreted over the rays employed in the proof, φ assumes the value 1, i.e., φ is identical with the unit of the algebra and is, therefore, "true" in the quantum logical sense of 'true.' But by the equivalence of 2-valued measures and 2-valued homomorphisms, φ is a classical *contradiction*. Hence, the very same notion of truth that, for the quantum logical

interpretation, validates excluded middle *also* validates classical contradictions. Thus, whatever the sense in which, on the quantum logical interpretation, every proposition is determinately true or false, it cannot be the classical one since it counts as true a proposition that is false under every classically possible truth value assignment to its propositional constituents. It remains an open problem for the interpretation to explain the notion of truth that it employs.

The intuitive picture that emerges on the view developed here takes the following form in the spin-1 case. For every orthogonal triple (x, y, z) of directions in physical space, a spin-1 particle carries with it an "instruction-set" for each of the three families of *basic* propositions $P_{x\theta}, P_{y\theta}, P_{z\theta}$. The instruction sets determine how the particle will answer any question regarding a proposition belonging to any such family. The sets are so constrained that properties mentioned in the basic propositions $P_{x\theta}$ and $P_{x\theta'}, \theta \neq \theta'$, occur with the same frequency. The classical or macroscopic level enters the analysis through the dependence of probability on the geometry of physical space. This is expressed by the broad empirical fact that the properties whose criteria of application sustain the simple geometrical relationship exhibited by $P_{x\theta}$ and $P_{x\theta'}, \theta \neq \theta'$ are statistically equivalent. The abstract representation of this situation is precisely what is given by the structure of the derived propositions.

The conceptually difficult step that the Hilbert space formalism embodies is the nature of its separation of the truth of distinct propositions such as $P_{x\theta}$ and $P_{x\theta'}$ from their probability. The novelty of the representation of the fact that such propositions can be statistically equivalent while differing in truth value consists in the inherent incompleteness the representation expresses. The interpretative difficulty the physical situation presents arises from the expectation—not fulfilled by spin-1 systems—that the properties that are constituent in basic propositions such as $P_{x\theta}$ and $P_{x\theta'}$ must be the *same* property, one which is merely indicated by different operational procedures.

The system $\boldsymbol{P}$ of basic propositions is a structure internal to the quantum theory itself, one that emerges as the basis for the theory's own hidden variable interpretation of its significance. What remains unusual from a pre-quantum mechanical perspective is the way in which properties are coupled with their measurement procedures; they—or rather, they and the propositions containing them—are detachable from their criteria of application only probabilistically. In answer to a question that was posed at the beginning of this study we can say that the *probability* of a proposition involving a directional property is independent of the orthogonal basis in E^3 which is selected by its associated ideal measurement procedure, but the *proposition itself* is not independent of such a basis. This may capture Pauli's (1994) contention that quantum mechanics rejects the idea of a "detached observer," since measurement remains opaque in the sense that the separation between a property and its criterion of application—familiar from our experience with classical physics—has been all but eliminated. We can achieve a separation of properties and their criteria of application only at the probabilistic level. This is the residual affinity with Heisenberg alluded to earlier in the context of his claim that the measurement process is intractable. But the affinity with Heisenberg is not being inexplicable by any possible theory, but is rather an internal analogue of intractability: Relative to the theoretical representation of basic elementary propositions, ideal operational procedures, and available

statistical information, our knowledge of basic propositions lacks a component that, like a classical mechanical state, is interpretable as knowledge of their truth value; what is expressible theoretically is knowledge only of their probability.

ACKNOWLEDGEMENT

The initial draft of this paper was completed during my tenure as a visiting fellow at All Souls College. My thanks to the College for providing me with such an excellent environment in which to pursue my work. I owe a special debt of gratitude to Jeremy Butterfield for numerous conversations related to the paper's composition. Itamar Pitowsky has been a constant and indispensable source of advice regarding all aspects of the ideas dealt with here. I wish to thank the Social Sciences and Humanities Research Council of Canada and the British Academy's International Collaborative Programmes for their financial support.

NOTES

[1] See DiSalle (2005) for an elaboration of this observation.

[2] The notion of a generalized probability measure was introduced by Gleason (1957) in the context of his characterization of the measures definable on the closed linear subspaces of Hilbert space. The analysis of the three-dimensional case proved to be fundamental. For this case, a *generalized probability measure* is a map f from the closed linear subspaces of H^3 to the closed unit interval satisfying the conditions

$$fa + fb \leq 1$$

for $a \perp b$, and

$$fa + fb + fc = 1$$

for any three rays a, b, c which are mutually orthogonal. The derived structure on which we have focused, namely that associated with the *square of the spin*$_x$, was specifically chosen for its isomorphic representation by a substructure of $B(H^3)$

[3] It was observed by Hultgren and Shimony (1977) that the spin propositions of a spin-1 system do not exhaust $B(H^3)$, so that the propositions of the derived structure form a substructure of the full three dimensional Hilbert space. This is in contrast with spin-1/2 propositions and $B(H^2)$. Shimony and Hultgren raised the question whether it is possible to give an operational motivation for the whole of $B(H^3)$. The question is answered positively in (Reck et al. 1994). Thanks to Pitowsky for bringing this to my attention and for the references just cited.

[4] For an extended discussion of Heisenberg see Frappier (2005).

REFERENCES

Bub, J. (1974). *The Interpretation of Quantum Mechanics* (Dordrecht: Reidel).

Bub, J. (1997). *Interpreting the Quantum World* (Cambridge: Cambridge University Press).

Demopoulos, W. (1977). "Completeness and realism in quantum mechanics," in R. Butts and J. Hintikka (eds.), *Foundational Problems in the Special Sciences* (Dordrecht: Reidel) 81–88.

Demopoulos, W. (2004). "Elementary propositions and essentially incomplete knowledge: a framework for the interpretation of quantum mechanics," *Noûs* **38** 86–109.

DiSalle, R. (2005). *Understanding Space-Time: The Philosophical Development of Physics from Newton to Einstein*, Cambridge: Cambridge University Press.

Frappier, M. (2005). *Heisenberg's Notion of Interpretation*, U.W.O. Ph.D. Thesis.

Friedman, M. and H. Putnam (1978). "Quantum logic, conditional probability and inference," *Dialectica* **32** 305–315.

Gleason, A.M. (1957). "Measures on the closed subspaces of Hilbert space," *Journal of Mathematics and Mechanics* **6** 885–893.

Heisenberg, W. (1927). "The physical content of quantum kinematics and mechanics," *Zeitschrift für Physik* **43** 172–198, reprinted in English translation in J. A. Wheeler and W. H. Zurek (eds.), *Quantum Theory and Measurement* (Princeton: Princeton University Press, 1983) 62–84.

Hultgren, B. O. and A. Shimony (1977). "The lattice of verifiable propositions of the spin-1 system," *Journal of Mathematical Physics* **18** 381–394.

Kochen, S. and E. P. Specker (1967). "The problem of hidden variables in quantum mechanics," *Journal of Mathematics and Mechanics* **17** 59–87.

Mackey, G. (1963). *Mathematical Foundations of Quantum Mechanics* (New York: Benjamin).

Pauli, W. (1994). *Writings on Physics and Philosophy*, Charles P. Enz and Karl von Meyenn (eds.), English translation by Robert Schlapp (Berlin: Springer-Verlaag).

Pitowsky, I. (1998). "Infinite and finite Gleason's theorem and the logic of indeterminacy," *Journal of Mathematical Physics* **39** 218–228.

Pitowsky, I. (2005). "Quantum mechanics as a theory of probability," *This Volume*.

Reck, M. et al. (1994). "Experimental realization of any discrete unitary operator," *Physical Review Letters* **73** 58–61.

von Neumann, J. (1932). *Mathematische Grundlagen der Quantenmechanik*, 1955 English translation by Robert Beyer (Princeton: Princeton University Press).

Wittgenstein, L. (1921). *Tractatus logico-philosophicus* (London: Kegan Paul).

Wittgenstein, L. (1929). "Some remarks on logical form," *Proceedings of the Aristotelian Society: Suppl. Vol. IX* 162–171.

CLARK GLYMOUR*

5. MARKOV PROPERTIES AND QUANTUM EXPERIMENTS

Few people have thought so hard about the nature of the quantum theory as has Jeff Bub, and so it seems appropriate to offer in his honor some reflections on that theory. My topic is an old one, the consistency of our microscopic theories with our macroscopic theories, my example, the Aspect experiments (Aspect et al., 1981, 1982, 1982a; Clauser and Shimony, 1978; Duncan and Kleinpoppen, 1998) is familiar, and my simplification of it is borrowed. All that is new here is a kind of diagonalization: an argument that the fundamental principles found to be violated by the quantum theory must be assumed to be true of the experimental apparatus used in the experiments that show the violation.

The chief principle I have in mind is essential in causal inference in macroscopic problems, and is used almost without notice in experimental and observational studies in economics, epidemiology, biology, physics, everywhere. The *Causal Markov Condition (CMC)* is the following property:

Consider any system S $= \langle G, Pr \rangle$, including a set $\boldsymbol{V}$ of variables whose causal relations are represented by a directed acyclic graph G having members of $\boldsymbol{V}$ as vertices. A directed edge, $V_1 \rightarrow V_2$ in G represents the proposition that there exists a set A of values for $\boldsymbol{V} \backslash \{V_1, V_2\}$ such that V_1 covaries with V_2 upon an intervention fixing $\boldsymbol{V} \backslash \{V_1, V_2\}$ and randomizing V_1.: Let $\boldsymbol{V}$ be *causally sufficient* : there is no variable X not in V such that if G were expanded to include X, there would be two vertices in V with edges from X directed into them. For any variable V in $\boldsymbol{V}$, let **Par** (V) be the set of vertices in $\boldsymbol{V}$ that have edges directed into V, and let **Des** (V) be the set of edges that are endpoints of directed paths from V. Let Pr be a joint probability distribution on all possible assignments of values to variables in $\boldsymbol{V}$ such that for all vertices V_1, V_2 in $\boldsymbol{V}$, and for all such assignments of values, if V_2 is not a member of **Des**(V_1), then V_1 is independent (in measure Pr) of V_2 conditional on **Par**(V_1). Then S satisfies the Causal Markov Condition.

Abstract as it may be, the condition is merely a reasonably rigorous generalization of Hans Reichenbach's "(1956) screening off" conditions for causal relations. Causally sufficient, feed-forward deterministic systems satisfy the condition if their exogenous causes are independent in probability.

* Dept. of Philosophy, Carnegie Mellon University, Pittsburgh PA 1521-3890, USA and Florida Institute for Human and Machine Cognition, University of West Florida, Pensacola, FL 32052, USA; E-mail: cg09@andrew.cmu.edu

W. Demopoulos and I. Pitowsky (eds.), Physical Theory and its Interpretation, 117–126.

A second principle is *Faithfulness* : All conditional independence relations in a system satisfying the Causal Markov Condition are consequences of that condition applied to the graph of the system.

One way to view the experiments that demonstrate the inconsistency of quantum theory with the Bell inequalities is that they show that one or both of these conditions must fail as universal causal principles: feed-forward systems exist that cannot be made causally sufficient consistent with CMC and Faithfulness. There are many diagnoses in different terms. David Bohm, Bub's teacher, would perhaps have said that that is because no system is causally sufficient; other commentators might locate the problem with the assumption of a joint probability distribution, and so on. I wish merely to point to the curiously valid, almost Wittgensteinian logic, that gets us to the inconsistency.

Instances of assumptions of the CMC and of Faithfulness could be traced through the details of the experimental set up, runs and data analyses of the Aspect experiments, But it has been a long time since I was any kind of physicist, and I would inevitably misrepresent details and confuse even the readers of clearest mind, and there are details of sensor behavior and sensitivity that complicate without clarifying. So I will pass on the details and consider instead a very simple idealization of the phenomenon, due to N. David Mermin (1985, 1990).

Consider two detectors I and II that are spatially separated. Each detector has three settings, $S = 1, 2$ or 3. Further each detector has a red bulb R and a green bulb G. Pairs of particles are emitted from a source and enter the two detectors. There is no other physical connection of any kind we know of between the detectors (Figure 5.1).

The detectors behave this way: (1) when both detectors are set to same value, no matter which, they both show red or they both show green. Red and green occur with equal frequency; (2) when the two detectors are set to any two *different* values, they show the same color, both red or both green, 1/4 of the time—again, red and green occur with equal frequency in this case, and different colors 3/4 of the time—each

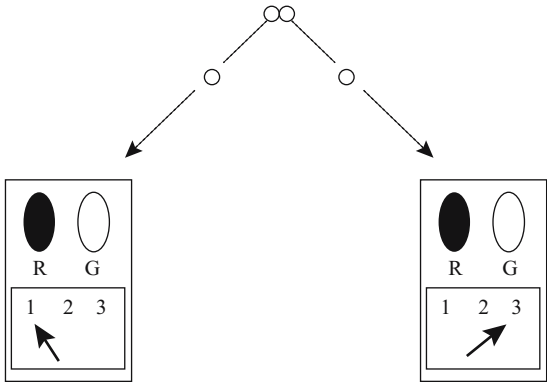

FIGURE 5.1.

Table 5.1

Left indicator setting	Left indicator light color	Right indicator setting	Right indicator light color	Probability of the two light colors given the settings
1	Red	1	Red	1
1	Green	1	Green	1
2	Red	2	Red	1
2	Green	2	Green	1
3	Red	3	Red	1
3	Green	3	Green	1
1	Red	2	Red	1/8
1	Green	2	Green	1/8
1	Red	2	Green	3/8
1	Green	2	Red	3/8
2	Red	1	Red	1/8
2	Green	1	Green	1/8
2	Red	1	Green	3/8
2	Green	1	Red	3/8
1	Red	3	Red	1/8
1	Green	3	Green	1/8
1	Red	3	Green	3/8
1	Green	3	Red	3/8
3	Red	1	Red	1/8
3	Green	1	Green	1/8
3	Red	1	Green	3/8
3	Green	1	Red	3/8
2	Red	3	Red	1/8
2	Green	3	Green	1/8
2	Red	3	Green	3/8
2	Green	3	Red	3/8
3	Red	2	Red	1/8
3	Green	2	Green	1/8
3	Red	2	Green	3/8
3	Green	2	Red	3/8

combination of colors (I green, II red; I red, II green) equally often. We can show the whole story about the probabilities with a tedious but clear table (Table 5.1).

The thing to notice immediately is that, no matter how we set the two detectors, the colors the detectors show will not be independent in probability. If both detectors are set at the same value, the probability that Detector II is red is 1 conditional on Detector I being red, and vice versa. If both detectors are set at different values, the probability that Detector II is green given that Detector I is red is three times the probability, on that same condition, that Detector II is red. Notice further, that someone at Detector I cannot use his settings of the detector to send signals or communications to someone at Detector II via the color that shows up at Detector II. For despite the fact that no matter how the detectors are set, the colors are correlated, the color at Detector II is independent in probability of the setting at Detector I.

Table 5.2

State	1,2	2,1	1,3	3,1	2,3	3,2
RRR	Same	Same	Same	Same	Same	Same
RRG	Same	Same	Differ	Differ	Differ	Differ
RGR	Differ	Differ	Same	Same	Differ	Differ
GRR	Differ	Differ	Differ	Differ	Same	Same
RGG	Differ	Differ	Differ	Differ	Same	Same
GRG	Differ	Differ	Same	Same	Differ	Differ
GGR	Same	Same	Differ	Differ	Differ	Differ
GGG	Same	Same	Same	Same	Same	Same

Mermin puts the problem this way. The only explanation (he says) for the first six rows of the probability table is that the particles each have internal states that specify their response to each state of a detector. The internal states of each particle specify what color it will activate for each of the three settings of the detector. Since there are 2 possible colors for each detector setting, and three settings, there are 8 possible internal states for each particle. If and only if (Mermin says) both particles have the same internal states will the colors of the two detectors agree when they have the same setting, for all 3 possible settings. So the states of the particles have to be perfectly correlated, the same. If one particle will make a detector go red on setting 1, red on setting 2, and green on setting 3, so will the other. So the question becomes: *is there a probability distribution over these possible internal states of the two particles that, consistent with their perfect correlation, agrees with probability table?* There is not. In particular, there is no way to assign probabilities to the particle states so that when the settings of the detectors are different, the detector colors agree less than 1/3 of the time. Let's do another table (Table 5.2). The columns indicate the settings of the two detectors when they are different, and the entries indicate for each state and pair of settings whether the colors of the detectors are the same or different.

In each row the fraction of cases in which the colors are the same is 1/3 or more. No matter what the relative frequency of the various particle states may be, if the detectors are set at any pair of distinct settings, the colors must be the same at least 1/3 of the time, but in the data for the experiment, for such settings the colors are the same only 1/4 of the time.

So what does this have to do with Markov Assumption and so forth? Two things. On the one hand, the conclusion of the example, while not inconsistent with the Markov Assumption, is inconsistent with the conjunction of the Markov Assumption and the claim that the state of the particle is the only causal connection between the detectors. On the other hand, while Mirmin's reasoning is perfectly correct, his argument depends on using the Markov Assumption. I will represent Mirmin's account of his experiment as a causal graph, like this (Figure 5.2).

The causal diagram and the Markov Assumption explain why the setting of Detector I cannot be used to send a signal to Detector II via the color that appears at Detector II—there is no causal pathway from Setting of Detector I to Color for

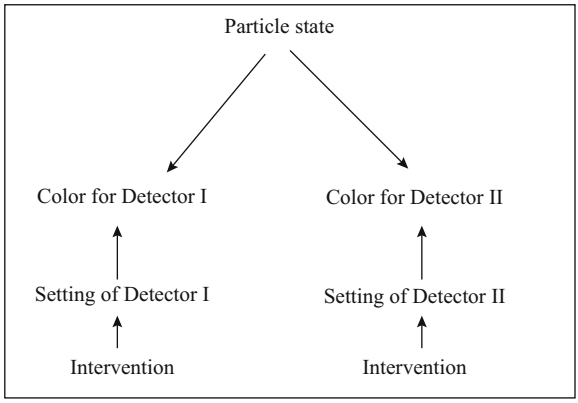

FIGURE 5.2.

Detector II, or vice-versa, so the two variables must be independent. And the causal diagram explains why the colors at the two detectors are correlated: they have a common cause. Nonetheless, there is something very wrong. There is no causal pathway from Color for Detector I to Color for Detector II, or in the other direction. There is no common cause of detector colors other than Particle State. Since Color for Detector II is not an effect of Color for Detector I, and vice versa, the Markov Assumption says they if the causal graph above is correct, the detector colors should be independent of one another *conditional* on Particle State. Indeed, that is exactly what Mermin's particle states do imply. For example, given that the particle state is RRR, then Detector I is red and Detector II is red: no matter the settings and neither detector provides any information about the other detector not already entailed by the particle state. If the particle state is RGR, then no matter how Detector I is set, the color in Detector I gives no further information about color that will appear at Detector II. (The setting chosen for Detector II provides further information about the color that will show up for Detector II when the particle is in the RGR state, but that is beside the point.) But Mirmin's argument shows that these particle states cannot be made consistent with the assumed observed frequencies of colors in each combination of settings shown in Table 5.1. So there are logically just three alternatives (1) Mirmin has sneaked in some extra assumption somewhere, or (2) the Markov Assumption is false for this case, or (3) there is no causal explanation of the correlations of the detector colors. Perhaps more than one of these alternatives is true.

Mirmin has certainly sneaked in some assumptions—all of them instances of the Markov Assumption–and the fact that he does not make them explicit may indicate that the Markov Assumption is so fundamental to our reasoning about experiments that we use it automatically, without notice. For *there is* a common cause explanation of the probabilities in Table 5.1. Here is the idea, first noted by Suppes and Zanotti (1981) in a more general case: Change the particle states so that they no longer just specify a color for each of the three settings of a detector. Now they specify a color

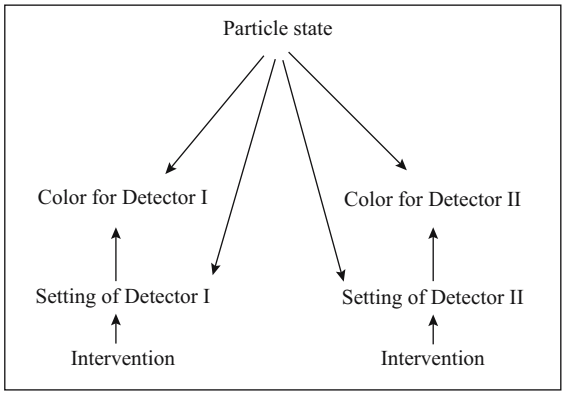

FIGURE 5.3.

for each detector setting *and a setting for each detector.* Instead of 8 internal states of the particle, we now have 48 internal states of the particle. The particle state now uniquely determines the color at each detector. Given the (new) particle state, the color at either detector provides no further information about the color at the other detector, because there is no more information to provide. We can give another causal diagram (Figure 5.3).

The Markov Assumption is satisfied. (Alternatively, the particle states can influence the interventions, which influence the detector settings.) Why doesn't Mirmin allow this? Because he thinks, quite reasonably, that the particle states do not cause the detector settings. Why not? Because he thinks the human act of setting the detectors (or a machine act of randomly setting the detectors is an *intervention,* a cause that is not influenced by any feature of the system and that fixes the value of the Detector setting while leaving all of the conditional probabilities of other variables unchanged. (Similar reasoning applies to the idea that the detector settings influence the particle state.)

Ok, take out the causal influence of the particle states on the detector settings, but leave the 48 states of the particle and their probabilities just as before:

Now we can still account for the correlations in Table 5.1, and the particle state is still a common cause of the detector colors, condition on which the detector colors are independent—the Markov Assumption is satisfied. Why doesn't Mirmin allow that? Because the causal diagram in Figure 5.4 and the probabilities assumed for the particle states are jointly inconsistent with the Markov assumption in another way—each detector setting is dependent in probability on the particle state (and vice-versa), but there is no causal pathway or common cause relating the detector setting variables to the particle state. Supposing there is another common cause beside the particle state that also influences the colors won't help things—the same argument goes through, its just more complicated. However, we do things, we do not have a causal explanation of the experiment consistent all the way through with the Markov assumption.

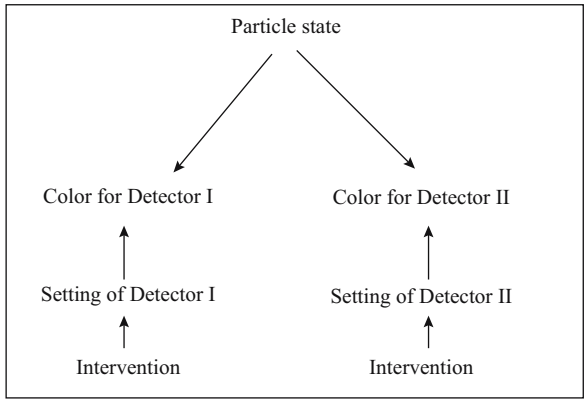

FIGURE 5.4.

Mirmin—and we—reason about his imaginary experiment using the Markov Assumption and the notion of an intervention, and yet the experiment allows of no causal explanation consistent with the Markov Assumption. The example is a simplification of what goes on in real experiments to test remote correlations predicted by a consequence of the quantum theory, Bell's theorem. In quantum experiments, we pull ourselves *down* by our bootstraps.

Now there is an obvious solution to the problem: the color at one or both of the detectors influences the color at the other detector.

This is a popular solution, and the reason why the problem is often said to be about "locality" or the phenomenon is said to exhibit "non-locality." Often the non-locality solution is implicitly motivated by the idea that the correlations between the colors must have a causal explanation.

Since the detectors can be far enough apart, and the color measurements close enough in time that the theory of relativity prohibits a signal from being sent from one detector to another, the solution has a problem. The problem is this: Suppose before the experiment, the guy at Detector II tells the guy at Detector I how Detector II will be set. Then, if the causal story above is correct, by adjusting the settings of Detector I the first guy can send signals to the second guy, who will figure them out from the color that shows up at Detector II. It works this way. There is in Figure 5.5 a causal pathway from setting of Detector I to the color at Detector II. The pathway must create an association between the two, and associations are all that is needed for communication, for sending a signal. The Faithfulness assumption says a direct causal connection creates an association—and the very point of the non-locality hypothesis is to create such an association between the colors. (Consistently with the Markov Assumption the association cannot be the effect of a common cause—for reasons we have already reviewed.) The setting of Detector I influences the color at Detector I, so we have a sequence of causal links—and correlations or associations—between Detector 1 and the color at Detector II. Now; a causal linkage of one variable with

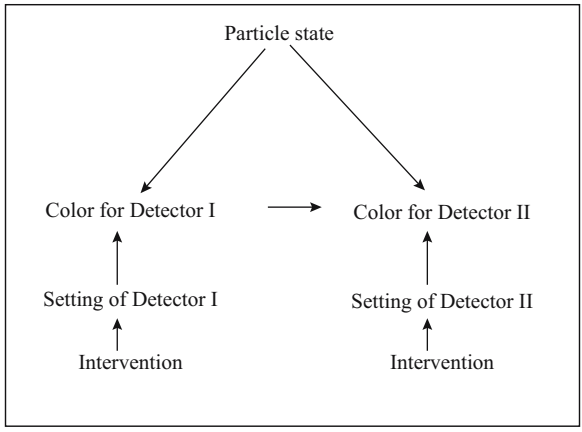

FIGURE 5.5.

a second linked with a third need not always create an association between the two variables, even if it is the only pathway connecting the variables (as in this case between Detector I and the color at Detector II). For example, suppose variable A has three values and variable B has three values (say, b1, b2 and b3) and variable C has two values, and the probabilities for two of the values (b1 and b2) of B depend on the value of A, (but the third value, b3, of B does not depend on the value of A) and the probability of values of C depends on whether B has value b3 or one of the values b1, b2, but doesn't depend on which of the values b1 or b2 B has. Then interventions that vary A will not create any association with C. Despite the fact that A influences B, and B influences C, A does not influence C: causation is not transitive. *But if B has only two values, the causal relations must be transitive, and A must be associated with C.* That is exactly the situation in the Mermin's thought experiment. Hence relativity can be violated. Having the influence go in both ways doesn't help; the argument still works.

The argument doesn't depend on any philosophical niceties about what "causation" means, and it doesn't depend on any details of the physics. It depends on the assumption that the settings of the Detectors are interventions, and the hypothesis that the "non-locality" relation creates an influence between the colors. So, if relativity is true and the statistics drawn from the Aspect and similar experiments are sound, causal non-locality is a non-starter.

The upshot is this: real experiments with associations analogous to those of Mermin's thought experiment create associations that have no causal explanation consistent with the Markov Assumption, and the Markov assumption must be applied, implicitly or explicitly, to obtain that conclusion. You can say that there is no causal explanation of the phenomenon, or that there is a causal explanation but it doesn't satisfy the Markov Assumption. I have no trouble with either alternative. It is not a truth of logic that all experimental associations have a causal explanation, and it is

not a truth of logic that all causal relations satisfy the Markov Assumption. That's up to Nature. But I do have this problem: *why, then, does the Markov Assumption work with our experiments on middle sized dry and wet goods, with climate, and rats and drugs, and so much else?*

I have no definite answer. I would suggest looking in these banal directions. First, among properties of middle sized objects, Aspect-like associations are extremely small, so the properties of systems are nearly deterministically related, or would be if all significant causes of variation were accounted for; second, when system are not causally sufficient, we make them nearly so when we can by redefining variables, by conditioning on variables with unexplained associations, and other devices; third, insofar as macroscopic frequencies are generated as "strike ratios" from deterministic processes, as proposed long ago by Hans Reichenbach in his doctoral thesis and more recently by Michael Strevens (2003), we should expect the Causal Markov Condition to hold necessarily. And finally, there are proofs that under continuous measures on the parameters of various families of probability distributions, the Markov Condition implies that the Faithfulness condition holds almost always (Spirtes et al., 2000).

REFERENCES

A. Aspect, P. Granger and G. Roger, *Physical Review Letters* 47, 460, 1981.
A. Aspect, P. G. Granger and G. Roger, *Physical Review Letters*, 49, 91, 1982.
A. Aspect, J. Daligard, and G. Roger, *Physical Review Letters*, 49, 1804, 1982a.
J. F. Clauser and A. Shimony, *Reports on Progress in Physics*, 41, 1881, 1978.
A. J. Duncan and H. Kleinpoppen, in *Quantum Mechanics versus Local Realism*, F. Selleri (ed.), Plenum, New York, 1998.
N. D. Mermin, *Physics Today* 38, 38–47, 1985.
N. D. Mermin, *Boojums All the Way Through*, Cambridge, Cambridge University Press, 1990.
P. Spirtes, C. Glymour, and R. Scheines, *Causation, Prediction and Search*, MIT Press, Cambridge, MA., 2000.
M. Strevens, *Bigger Than Chaos*, Cambridge, MA Harvard University Press, 2003.
P. Suppes and M. Zanotti, When are probabilistic explanations possible?, *Synthese* 48, 191, 1981.

STAN GUDDER*

6. QUANTUM ENTROPY

The entropy concept has proved to be very useful in the fields of information theory and the statistics of data processing [1, 2]. Entropy in conjunction with information – theoretic and combinatorial methods has also been applied to derive many well-known inequalities [3]. More recently quantum entropy has become important in quantum information theory [2, 4]. The derived concept of relative entropy is also useful because it provides a measure of the distance between two probability distributions or between two quantum states. Our main concerns here will be the proofs of two basic properties of quantum relative entropy, namely positivity and monotonicity. It turns out that the proof of positivity is easy while the proof of monotonicity is difficult. For an arbitrary Hilbert space, monotonicity and the closely related property of strong subadditivity were open problems for a number of years until they were solved by Uhlmann [5] and by Lieb and Ruskai [6]. If the Hilbert space is finite-dimensional then a much simpler approach due to Petz [2, 4] can be taken. Our main contribution will be some clarification of this approach. The finite-dimensional case is still important because it is the basic arena for quantum computation and information theory [2, 7]. We believe that this work provides a beautiful application of the techniques of linear algebra.

1 CLASSICAL ENTROPY

Before we tackle quantum entropy, let us warm up with a brief discussion of classical entropy. Let $\Omega = \{\omega_1, \ldots, \omega_n\}$ be a finite sample space and let $P(\omega_i) = p_i$ be a probability distribution on Ω. Then p_i is the probability that the outcome ω_i occurs and we have that $p_i \geq 0$, $\sum p_i = 1$. The *Shannon entropy* $S(P)$ is defined by

$$S(P) = -\sum p_i \ln p_i \tag{1.1}$$

For (1.1) to make sense when $p_i = 0$ we define $0 \ln 0 = 0$. Now $S(P) \geq 0$ provides the lack of information (or ignorance or uncertainty) about our statistical system given by the distribution P. In other words, $S(P)$ is a measure of the unpredictability encoded in P that a particular outcome will occur. In the case of maximal ignorance we cannot predict at all which outcome will occur so we have the uniform distribution $p_i = 1/n$,

* Department of Mathematics, University of Denver, Denver, Colorado 80208, stan.gudder@nsm.du.edu

W. Demopoulos and I. Pitowsky (eds.), Physical Theory and its Interpretation, 127–142.

$i = 1, \ldots, n$. In this case

$$S(P) = -\sum_{i=1}^{n} \frac{1}{n} \ln \frac{1}{n} = -\ln \frac{1}{n} = \ln n$$

We shall show shortly that the value $S(P) = \ln n$ is the maximal entropy for Ω. At the other extreme, if we have complete information about the system, then we can predict exactly which outcome will occur. We then have that $p_i = 1$ for some $1 \leq i \leq n$. Hence, $S(P) = 0$ which is the minimal value for the entropy. It is clear that these are the only kinds of distributions that attain the minimal value.

Now suppose we have two probability distributions $P(\omega_i) = p_i$ and $Q(\omega_i) = q_i$ on the sample space $\Omega = \{\omega_1, \ldots, \omega_n\}$. We say that P is *absolutely continuous* relative to Q and write $P \prec Q$ if $q_i = 0$ implies that $p_i = 0$, $i = 1, \ldots, n$. The *relative entropy* of P with respect to Q is defined by

$$S(P \mid Q) = \sum (p_i \ln p_i - p_i \ln q_i) = -S(P) - \sum p_i \ln q_i$$

if $P \prec Q$ and $S(P \mid Q) = \infty$ otherwise. We may think of $S(P \mid Q)$ as a measure of the distance between P and Q. Unfortunately, $S(P \mid Q) \neq S(Q \mid P)$ in general. For example, if Q is the uniform distribution on Ω and P is the exact distribution $P(\omega_i) = 1$ then $P \prec Q$ and $S(P \mid Q) = \ln n$ but $Q \not\prec P$ so that $S(Q \mid P) = \infty$. However, our first theorem shows that relative entropy possesses the important property of distance called *strict positivity*.

Theorem 1.1 *Relative entropy satisfies* $S(P \mid Q) \geq 0$ *with* $S(P \mid Q) = 0$ *if and only if* $P = Q$.

Proof. We may assume without loss of generality that $p_i > 0$, $i = 1, \ldots, n$. If $P \not\prec Q$, then $S(P \mid Q) = \infty > 0$ so assume that $P \prec Q$ in which case $q_i > 0$, $i = 1, \ldots, n$. Applying the well-known calculus inequality $\ln x \leq x - 1$ for $x > 0$ we have that

$$S(P \mid Q) = -\sum p_i \ln \frac{q_i}{p_i} \geq \sum p_i \left(1 - \frac{q_i}{p_i}\right) = \sum (p_i - q_i) = 0$$

This proves positivity and we now prove strictness. It is clear that $P = Q$ implies that $S(P \mid Q) = 0$ so suppose that $P \neq Q$. Then there exist p_j, q_j such that $p_j \neq q_j$. Again, if $P \not\prec Q$ then $S(P \mid Q) = \infty \neq 0$ so assume that $P \prec Q$. Now we see from the graphs that $\ln x = x - 1$ only at the point $x = 1$. It follows that

$$-p_j \ln \frac{q_j}{p_j} > p_j \left(1 - \frac{q_j}{p_j}\right)$$

Hence,

$$S(P, Q) > \sum p_i \left(1 - \frac{q_i}{p_i}\right) = 0$$

□

As an application of Theorem 1.1, let Q be the uniform distribution on Ω. Then for any distribution P on Ω we have that

$$0 \leq S(P \mid Q) = -S(P) - \sum p_i \ln \frac{1}{n} = -S(P) + \ln n$$

Hence, $S(P) \leq \ln n$ so the uniform distribution has maximal entropy and is the unique distribution with this property.

We now discuss another important property of relative entropy called monotonicity. Suppose we have two finite sample spaces $\Omega_1 = \{\omega_1^1, \ldots, \omega_m^1\}$, $\Omega_2 = \{\omega_1^2, \ldots, \omega_n^2\}$ and we form the joint sample space

$$\Omega_1 \times \Omega_2 = \left\{(\omega_i^1, \omega_j^2) \colon i = 1, \ldots, m, j = 1, \ldots n\right\}$$

Let $P_{12}(\omega_i^1, \omega_j^2) = p_{ij}$ be a joint probability distribution so that $p_{ij} \geq 0$, $\sum p_{ij} = 1$. Then p_{ij} gives the probability that outcome ω_i^1 occurs in the first system and outcome ω_j^2 occurs in the second system. The *marginal distributions* are given by $P_1(\omega_i^1) = p_i^1$ where $p_i^1 = \sum_j p_{ij}$ and $P_2(\omega_j^2) = p_j^2$ where $p_j^2 = \sum_i p_{ij}$. We interpret P_1 as the distribution on system 1 when system 2 is disregarded and a similar interpretation is given for P_2. Monotonicity says that if P_{12} and Q_{12} are joint distributions on $\Omega_1 \times \Omega_2$ then

$$S(P_1 \mid Q_1) \leq S(P_{12} \mid Q_{12}) \tag{1.2}$$

Thus, if we disregard system 2 the relative entropy cannot increase. In other words, a joint system can distinguish two distributions better than a single system can distinguish their marginal distributions.

Theorem 1.2 *The monotonicity property (1.2) holds.*

Proof. If $P_{12} \not\prec Q_{12}$ we are finished so assume that $P_{12} \prec Q_{12}$. It easily follows that $P_1 \prec Q_1$. We can then assume without loss of generality that $p_{i,j}, q_{ij}, p_i^1, q_i^1$ are all positive. To prove (1.2) we first write it as

$$\sum_j p_j^1 (\ln p_j^1 - \ln q_j^1) \leq \sum_{j,k} p_{jk} (\ln p_{jk} - \ln q_{jk}) \tag{1.3}$$

Now (1.3) is equivalent to

$$\sum_{j,k} p_{jk} \ln \frac{p_j^1}{q_j^1} \leq \sum_{j,k} p_{jk} \ln \frac{p_{jk}}{q_{jk}}$$

which can be rewritten as

$$\sum_{j,k} p_{jk} \ln \frac{p_j^1 q_{jk}}{q_j^1 p_{jk}} \leq 0 \tag{1.4}$$

To prove (1.4) apply the inequality $\ln x \leq x - 1$ to obtain

$$\sum_{j,k} p_{jk} \ln \frac{p_j^1 q_{jk}}{q_j^1 p_{jk}} \leq \sum_{j,k} p_{jk} \left(\frac{p_j^1 q_{jk}}{q_j^1 p_{jk}} - 1 \right) = \sum_{j,k} \left(\frac{p_j^1 q_{jk}}{q_j^1} - p_{jk} \right)$$

$$= \sum_{j,k} \frac{p_j^1 q_{jk}}{q_j^1} - 1 = \sum_j p_j^1 - 1 = 0$$

□

2 OPERATOR CONVEXITY

Before we can study quantum entropy we need some background in linear algebra, in particular matrix theory. For $n \in \mathbb{N}$, let V be an n-dimensional complex inner product space with inner product $\langle \psi, \phi \rangle$. Now V is isomorphic to the inner product space $\mathbb{C}^n$ with the usual inner product

$$\langle \psi, \phi \rangle = \sum_{i=1}^n \alpha_i \overline{\beta}_i$$

where $\psi = (\alpha_1, \ldots, \alpha_n)$, $\phi = (\beta_1, \ldots, \beta_n)$. For this reason we shall usually assume that $V = \mathbb{C}^n$. Denoting the set of linear operators on V by $\mathcal{L}(V)$, any $A \in \mathcal{L}(V)$ can be represented by a matrix operator on $\mathbb{C}^n$. Again, we shall usually assume that $A \in M_n$ where M_n is the set of $n \times n$ complex matrices.

Let S_n be the set of hermitian $n \times n$ complex matrices and let I_n be the identity matrix. The spectral theorem states that any $A \in S_n$ has the form $A = \sum_{i=1}^n \lambda_i P_i$ where $\lambda_i \in \mathbb{R}$ are the eigenvalues of A and P_i are one-dimensional orthogonal projections satisfying $P_i P_j = 0$, $i \neq j$, and $\sum P_i = I_n$. Equivalently, there exists a diagonal matrix $D = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$, $\lambda_i \in \mathbb{R}$, and a unitary matrix U such that $A = UDU^*$. If $f : \mathbb{R} \to \mathbb{R}$ and $A \in S_n$ we define $f(A)$ by

$$f(A) = \sum_{i=1}^n f(\lambda_i) P_i$$

or equivalently $f(A) = Uf(D)U^*$ where $f(D) = \operatorname{diag}(f(\lambda_1), \ldots, f(\lambda_n))$. Notice that if $g(\lambda) = f(\lambda)$ for all $\lambda \in \sigma(A) = \{\lambda_i : 1 \leq i \leq n\}$, then $g(A) = f(A)$. In particular, there exists a polynomial $p(x) = \sum c_i x^i$ such that $p(A) = f(A)$. Hence, we can write

$$f(A) = p(A) = \sum c_i A^i \tag{2.1}$$

We can apply (2.1) to obtain a result that we shall find useful. Suppose $A, B \in S_n$ with $AB = 0$. By taking adjoints of both sides we obtain $BA = 0$. If f satisfies $f(0) = 0$, then $c_0 = 0$ and (2.1) gives

$$f(A + B) = \sum c_i (A + B)^i = \sum c_i (A^i + B^i) = f(A) + f(B) \tag{2.2}$$

We say that $A \in S_n$ is *positive* and write $A \geq 0$ if $\langle A\psi, \psi\rangle \geq 0$ for all $\psi \in \mathbb{C}^n$. Also, $A \in S_n$ is *strictly positive* and we write $A > 0$ if $\langle A\psi, \psi\rangle > 0$ for all $\psi \in \mathbb{C}^n$ with $\psi \neq 0$. It is easy to show that $A \in S_n$ is (strictly) positive if and only if $(\lambda > 0)\lambda \geq 0$ for all $\lambda \in \sigma(A)$. Also, $A \in S_n$ is strictly positive if and only if A is invertible and $A \geq 0$. We denote the set of positive matrices in M_n by S_n^+. For $A, B \in S_n$ we define $A \leq B$ if $B - A \geq 0$.

We say that a function $f: (0, \infty) \to \mathbb{R}$ is *operator convex* [8] if for every $\lambda \in [0, 1]$ and every $A > 0$, $B > 0$ we have

$$f(\lambda A + (1 - \lambda)B) \leq \lambda f(A) + (1 - \lambda) f(B)$$

This notion generalizes the concept of convex functions in the ordinary sense. In calculus courses, convex functions are called concave upward and twice differentiable concave upward functions are characterized by $f''(x) \geq 0$. Two examples of convex functions are $f(x) = 1/x$ and $g(x) = -\ln x$. It turns out that a convex function need not be operator convex [8]. However, we shall show that $f(x) = 1/x$ and $g(x) = -\ln x$ are operator convex. But first, let us look at some examples. The following computation shows that $f(x) = x^2$ is operator convex. Letting $\lambda \in [0, 1]$, $A, B > 0$ we have that

$$\begin{aligned}\lambda A^2 + (1 - \lambda)B^2 - [\lambda A + (1 - \lambda)B]^2 &= \lambda(1 - \lambda)(A^2 + B^2 - AB - BA) \\ &= \lambda(1 - \lambda)(A - B)^2 \geq 0\end{aligned}$$

Although this result is not very surprising, surprising things can happen even with a simple function like $f(x) = x^2$. Even though $f(x) = x^2$ is increasing on $(0, \infty)$, f is not operator increasing. That is, $0 \leq A \leq B$ does not imply that $A^2 \leq B^2$. To show this we use the well-known fact that $A \in M_2$ is positive if and only if the diagonal elements and determinant of A are nonnegative. Letting

$$A = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}, \qquad B = \begin{bmatrix} 3 & 1 \\ 1 & 1 \end{bmatrix}$$

it follows that $0 < A \leq B$. However,

$$B^2 - A^2 = \begin{bmatrix} 5 & 1 \\ 1 & 0 \end{bmatrix} \not\geq 0$$

so that $A^2 \not\leq B^2$. We next show that $f(x) = x^3$ is not operator convex. Letting $A, B \in M_2$ be defined as before, we have

$$\frac{1}{2}(A^3 + B^3) - \frac{1}{8}(A + B)^3 = \begin{bmatrix} 1.875 & 0.25 \\ 0.25 & 0 \end{bmatrix} \not\geq 0$$

Lemma 2.1 *The function $g(x) = -\ln x$ is operator convex.*

Proof. We first show that $f(x) = 1/x$ is operator convex. Note that if $A \leq B$ then $CAC^* \leq CBC^*$ for every $C \in M_n$. Indeed, we have that

$$\langle CAC^*\psi, \psi\rangle = \langle AC^*\psi, C^*\psi\rangle \leq \langle BC^*\psi, C^*\psi\rangle = \langle CBC^*\psi, \psi\rangle$$

for every $\psi \in \mathbb{C}^n$. Since $f(x) = 1/x$ is a convex function we have that

$$[\lambda x + (1-\lambda)y]^{-1} \leq \lambda x^{-1} + (1-\lambda)y^{-1}$$

for every $x, y \in (0, \infty)$, $\lambda \in [0, 1]$. Now let $A > 0$ and $B > 0$. Since I and A commute, they are simultaneously diagonalizable. It follows that

$$[\lambda I + (1-\lambda)A]^{-1} \leq \lambda I + (1-\lambda)A^{-1} \tag{2.3}$$

Applying (2.3) we have that

$$\begin{aligned}\left[\lambda I + (1-\lambda)A^{-1/2}BA^{-1/2}\right]^{-1} &\leq \lambda I + (1-\lambda)(A^{-1/2}BA^{-1/2})^{-1}\\ &= \lambda I + (1-\lambda)A^{1/2}B^{-1}A^{1/2}\end{aligned}$$

Hence,

$$\begin{aligned}[\lambda A + (1-\lambda)B]^{-1} &= \left[A^{1/2}\left(\lambda I + (1-\lambda)A^{-1/2}BA^{-1/2}\right)A^{1/2}\right]^{-1}\\ &= A^{-1/2}\left[\lambda I + (1-\lambda)A^{-1/2}BA^{-1/2}\right]^{-1}A^{-1/2}\\ &\leq A^{-1/2}\left[\lambda I + (1-\lambda)A^{1/2}B^{-1}A^{1/2}\right]A^{-1/2}\\ &= \lambda A^{-1} + (1-\lambda)B^{-1}\end{aligned}$$

To show that $g(x) = -\ln x$ is operator convex we employ the representation

$$-\ln x = \int_0^\infty \left(\frac{1}{x+t} - \frac{1}{1+t}\right)dt$$

from which we obtain for $A > 0$ that

$$-\ln A = \int_0^\infty \left[(A+tI)^{-1} - (1-t)^{-1}I\right]dt \tag{2.4}$$

By the operation convexity of $f(x) = 1/x$ we have that

$$\begin{aligned}[\lambda A + (1-\lambda)B + tI]^{-1} &= [\lambda(A+tI) + (1-\lambda)(B+tI)]^{-1}\\ &\leq \lambda(A+tI)^{-1} + (1-\lambda)(B+\lambda I)^{-1}\end{aligned} \tag{2.5}$$

Applying (2.4) and (2.5) gives

$$
\begin{aligned}
&-\ln(\lambda A+(1-\lambda)B)\\
&\quad=\int_0^\infty\Big[(\lambda A+(1-\lambda)B+tI)^{-1}-(1+t)^{-1}I\Big]\,dt\\
&\quad\le\int_0^\infty\Big[\lambda(A+tI)^{-1}+(1-\lambda)(B+tI)^{-1}-(1+t)^{-1}I\Big]\,dt\\
&\quad=\int_0^\infty\Big[\lambda(A+tI)^{-1}=\lambda(1+t)^{-1}I\Big]\,dt\\
&\qquad+\int_0^\infty\Big[(1-\lambda)(B+tI)^{-1}-(1-\lambda)(1_t)^{-1}I\Big]\,dt\\
&\quad=-\lambda\ln A-(1-\lambda)\ln B
\end{aligned}
$$

□

A linear transformation $U:\mathbb{C}^n\to\mathbb{C}^m$ is called an *isometry* if $U^*U=I_n$. It follows from a linear algebra result that if an isometry U is surjective, then U is unitary, that is, $UU^*=I_m$. Notice that if U is unitary and $f:\mathbb{R}\to\mathbb{R}$ then $f(U^*AU)=U^*f(A)U$ for every $A\in S_n$. Indeed, applying the spectral representation $A=\sum\lambda_iP_i$ we have that

$$
\begin{aligned}
f(U^*AU)&=f\left(\sum\lambda_iU^*P_iU\right)=\sum f(\lambda_i)U^*P_iU\\
&=U^*\sum f(\lambda_i)P_iU=U^*f(A)U
\end{aligned}
$$

Lemma 2.2 *If $f:(0,\infty)\to\mathbb{R}$ is operator convex and $U:\mathbb{C}^n\to\mathbb{C}^m$ is an isometry, then $f(U^*AU)\le U^*f(A)U$ for all $A\in S_m$ with $A>0$.*

Proof. Since $U^*AU>0$ when $A>0$ we can extend f to $[0,\infty)$ with $f(0)=0$ and nothing will change. To simplify the notation let $V=\mathbb{C}^n$, $W=\mathbb{C}^m$ and let W' be the range of U which is a subspace of W. Let $P:W\to W'$ be the projection onto W' and let $Q=I-P$ be the projection onto the orthocomplement of W'. Since $PU=U$ is a unitary transformation from V to W' and since PAP may be regarded as a matrix acting on W' we have that

$$
\begin{aligned}
f(U^*AU)&=f\left(U^*P(PAP)PU\right)=U^*Pf(PAP)PU\\
&=U^*f(PAP)U
\end{aligned}
\tag{2.6}
$$

If we can show that

$$
f(PAP)\le Pf(A)P \tag{2.7}
$$

then it would follow that

$$
f(U^*AU)\le U^*Pf(A)PU=U^*f(A)U
$$

which is our result. To prove (2.7) note that by (2.2) we have

$$f(PAP + QAQ) = f(PAP) + f(QAQ)$$

and $Pf(QAQ)P = 0$. It follows that

$$f(PAP) = Pf(PAP)P = Pf(PAP + QAQ)P \tag{2.8}$$

Defining the operator

$$S = P - Q = 2P - I$$

on W we see that $SS^* = S^*S = I$ so S is unitary. Now

$$\begin{aligned} \frac{A + S^*AS}{2} &= \frac{(P+Q)A(P+Q) + (P-Q)A(P-Q)}{2} \\ &= PAP + QAQ \end{aligned} \tag{2.9}$$

Applying the operator convexity of f and using (2.9) twice we have that

$$\begin{aligned} f(PAP + QAQ) &\leq \frac{1}{2}\left[f(A) + f(S^*AS)\right] = \frac{1}{2}\left[f(A) + S^*f(A)S\right] \\ &= Pf(A)P + Qf(A)Q \end{aligned} \tag{2.10}$$

Finally, by (2.8) and (2.10) we have that

$$f(PAP) = Pf(PAP + QAQ)P \leq Pf(A)P$$

which is (2.7). □

Finally, we need to recall that the *trace* of a square matrix $A = [a_{ij}]$ is given by $\mathrm{tr}(A) = \sum a_{ii}$, that is, the sum of the diagonal terms. Equivalently, if $A \in M_n$ and ψ_i is an orthonormal basis for $\mathbb{C}^n$ then

$$\mathrm{tr}(A) = \sum \langle A\psi_i, \psi_i \rangle$$

A standard property of the trace is that $\mathrm{tr}(AB) = \mathrm{tr}(BA)$ for all $A, B \in M_n$.

3 QUANTUM ENTROPY

We now consider the quantum generalizations of entropy and relative entropy. The quantum counterpart of a probability distribution is a *density matrix* (or *statistical matrix*) $\rho \in S_n^+$ with $\mathrm{tr}(\rho) = 1$. One reason for this correspondence is that the density matrices are precisely the elements of S_n whose eigenvalues form a probability distribution. We call a density matrix a *state* and denote the set of states on $\mathbb{C}^n$ by D_n.

The *von Neumann entropy* of $\rho \in D_n$ is defined by $S(\rho) = -\mathrm{tr}(\rho \ln \rho)$. Analogous to the classical situation $S(\rho) \geq 0$, S has its minimal value $S(P) = 0$ if P is a one-dimensional projection (these are called *pure states*) and S has its maximal value $S(I_n/n) = \ln n$ on the completely mixed state I/n.

For $\rho, \sigma \in D_n$ we write $\rho \prec \sigma$ if their null spaces satisfy $\mathrm{Null}(\sigma) \subseteq \mathrm{Null}(\rho)$. The *quantum relative entropy* of ρ with respect to σ is defined by

$$S(\rho \mid \sigma) = \mathrm{tr}(\rho \ln \rho - \rho \ln \sigma) = -S(\rho) - \mathrm{tr}(\rho \ln \sigma)$$

if $\rho \prec \sigma$ and $S(\rho \mid \sigma) = \infty$ otherwise. The next result is the quantum counterpart of Theorem 1.1.

Theorem 3.1 *The quantum relative entropy satisfies* $S(\rho \mid \sigma) \geq 0$ *with equality if and only if* $\rho = \sigma$.

Proof. As in the proof of Theorem 1.1 we can assume that $\rho > 0$ and $\rho \prec \sigma$. Let $\rho = \sum p_i P_i$ and $\sigma = \sum q_j Q_j$ be the spectral representations of ρ and σ where $p_i, q_i > 0$ with $\sum p_i = \sum q_j = 1$. Evaluating the trace using an orthonormal basis of eigenvectors for ρ gives

$$\begin{aligned} S(\rho \mid \sigma) &= \sum \langle (\rho \ln \rho - \rho \ln \sigma)\psi_i, \psi_i \rangle \\ &= \sum p_i \ln p_i - \sum \langle \rho \ln \sigma \psi_i, \psi_i \rangle \\ &= \sum p_i \ln p_i - \sum p_i \langle \ln \sigma \psi_i, \psi_i \rangle \end{aligned}$$

Now

$$\langle \ln \sigma \psi_i, \psi_i \rangle = \left\langle \sum_j (\ln q_j) Q_j \psi_i, \psi_i \right\rangle = \sum_j p_{ij} \ln q_j$$

where $p_{ij} = \langle Q_j \psi_i, \psi_i \rangle \geq 0$. Hence,

$$S(\rho \mid \sigma) = \sum_i p_i \left(\ln p_i - \sum_j p_{ij} \ln q_j \right)$$

Notice that

$$\sum_i p_{ij} = \mathrm{tr}(Q_j) = 1$$

and

$$\sum_j p_{ij} = \left\langle \sum_j Q_j \psi_i, \psi_i \right\rangle = \langle \psi_i, \psi_i \rangle = 1$$

so $[p_{ij}]$ is a doubly stochastic matrix. Letting $r_i = \sum_j p_{ij} q_j$, by the convexity of $-\ln x$ we have that

$$\sum_j p_{ij} \ln q_j \leq \ln \left(\sum_j p_{ij} q_j \right) = \ln r_i$$

with equality if and only if there exists a j such that $p_{ij} = 1$. Hence,

$$S(\rho \mid \sigma) \geq \sum p_i (\ln p_i - \ln r_i) = \sum p_i \ln \frac{p_i}{r_i} \tag{3.1}$$

with equality if and only if for every i there exists a j such that $p_{ij} = 1$; that is, if and only if $[p_{ij}]$ is a permutation matrix. Now the right hand side of (3.1) has the form of a classical relative entropy. It follows from Theorem 1.1 that $S(\rho \mid \sigma) \geq 0$ with equality if and only if $p_i = r_i$ for every i and $[p_{ij}]$ is a permutation matrix. If $S(\rho \mid \sigma) = 0$, we can relabel the basis of eigenvectors of ρ if necessary so that $[p_{ij}]$ is the identity matrix. It follows that $Q_i = P_i$, $i = 1, \ldots, n$. Moreover, $p_i = r_i = q_i$, $i = 1, \ldots, n$, so that $\rho = \sigma$. □

We would now like to obtain a monotonicity inequality analogous to (1.2) for quantum relative entropy. But first we need to understand the concept of a joint quantum system under a joint state. If V and W are finite-dimensional inner product spaces, their *tensor product* $V \otimes W$ can be thought of as the set of elements of the form $\sum_{i,j=1}^{n} v_i \otimes w_j$, $v_i \in V$, $w_j \in W$ where $v \otimes w$ satisfies
(1) $v \otimes (w_1 + w_2) = v \otimes w_1 + v \otimes w_2$
(2) $(v_1 + v_2) \otimes w = v_1 \otimes w + v_2 \otimes w$
(3) $\alpha(v \otimes w) = (\alpha v) \otimes w = v \otimes (\alpha w)$ for all $\alpha \in \mathbb{C}$.
Then $V \otimes W$ becomes an inner product space if we define

$$\langle v_1 \otimes w_1, v_2 \otimes w_2 \rangle = \langle v_1, v_2 \rangle \langle w_1, w_2 \rangle$$

and extend by linearity. The linear operators on $V \otimes W$ all have the form $\sum_{i,j=1}^{n} A_i \otimes B_j$, $A_i \in \mathcal{L}(V)$, $B_j \in \mathcal{L}(W)$ where

$$(A \otimes B)(v \otimes w) = Av \otimes Bw$$

and again we extend by linearity. It is easy to show that if v_i, w_j are orthonormal bases for V and W, respectively, then $v_i \otimes w_j$, $i = 1, \ldots, n, j = 1, \ldots, m$ is an orthonormal basis for $V \otimes W$. It follows that $\mathbb{C}^n \otimes \mathbb{C}^m$ is isomorphic to $\mathbb{C}^{nm}$ and $\mathcal{L}(\mathbb{C}^n \otimes \mathbb{C}^m)$ is isomorphic to M_{nm}.

If V_1 and V_2 correspond to two quantum systems, then the joint (or compound or composite) system for the two corresponds to $V_1 \otimes V_2$. Moreover, the joint states of the compound system are represented by density operators on $V_1 \otimes V_2$. We can assume that $V_1 = \mathbb{C}^n$, $V_2 = \mathbb{C}^m$ so that the joint system corresponds to $\mathbb{C}^n \otimes \mathbb{C}^m \approx \mathbb{C}^{nm}$ and the set of joint states corresponds to D_{nm}. Let M be a matrix for the compound

system so that M is a $nm \times nm$ matrix. If M has the form $A \otimes B$ we define the *partial trace* over the second system tr_2 by

$$\mathrm{tr}_2(M) = \mathrm{tr}_2(A \otimes B) = \mathrm{tr}(B)A$$

We then extend this definition by linearity. That is,

$$\mathrm{tr}_2\left(\sum A_i \otimes B_j\right) = \sum \mathrm{tr}(B_j)A_i$$

We define the partial trace over the first system tr_1 in a similar way. Notice that for $M = A \otimes B$ and orthonormal bases v_i, w_j we have that

$$\begin{aligned}\mathrm{tr}(M) = \sum \langle A \otimes B(v_i \otimes w_j), v_i \otimes w_j\rangle &= \sum \langle Av_i \otimes Bw_j, v_i \otimes w_j\rangle \\ &= \sum \langle Av_i, v_i\rangle\langle Bw_j, w_j\rangle \\ &= \mathrm{tr}(A)\mathrm{tr}(B) = \mathrm{tr}_1\left[\mathrm{tr}_2(A \otimes B)\right] \\ &= \mathrm{tr}_1\left(\mathrm{tr}_2(M)\right)\end{aligned}$$

It follows by linearity that $\mathrm{tr}_1(\mathrm{tr}_2(M)) = \mathrm{tr}(M)$ holds for any $nm \times nm$ matrix. If $\rho_{12} \in D_{nm}$ is a joint density operator we define the corresponding *marginal states* by $\rho_1 = \mathrm{tr}_2(\rho_{12})$ and $\rho_2 = \mathrm{tr}_1(\rho_{12})$.

In order to prove the monotonicity inequality for quantum relative entropy we consider M_n as a linear space with inner product $\langle A, B\rangle = \mathrm{tr}(AB^*)$. Since the dimension of M_n as an inner product space is n^2, M_n is isomorphic to $\mathbb{C}^{n^2}$. For $\sigma \in D_n$ $\sigma > 0$, we define the superoperators (linear operators on matrices) L_σ, R_σ by $L_\sigma(A) = \sigma A$, $R_\sigma(A) = A\sigma^{-1}$. It is easy to show that $L_\sigma \geq 0$. Indeed for every $A \in M_n$ since $A^*\sigma A \geq 0$ we have that

$$\langle L_\sigma(A), A\rangle = \langle \sigma A, A\rangle = \mathrm{tr}(\sigma AA^*) = \mathrm{tr}(A^*\sigma A) \geq 0$$

In a similar way, $R_\sigma \geq 0$. For $\sigma, \rho \in D_n$ we define the *relative modular operator* $\Delta(\sigma, \rho)$ by $\Delta(\sigma, \rho) = L_\sigma R_\rho$. Since L_σ and R_ρ commute, it follows that $\Delta(\sigma, \rho) \geq 0$.

For $\sigma \in D_n$, $\sigma > 0$ there exists a polynomial $p(x) = \sum c_i x^i$ such that $p(\sigma) = \ln \sigma$ and $p(L_\sigma) = \ln L_\sigma$. Hence, for any $A \in M_n$ we have that

$$\ln(L_\sigma)(A) = \sum c_i L_\sigma^i(A) = \sum c_i \sigma^i A = \ln(\sigma)A$$

In a similar way we have that $\ln(R_\sigma)(A) = -A\ln(\sigma)$. Moreover, since L_σ and R_ρ commute it follows that

$$\ln \Delta(\sigma, \rho) = \ln L_\sigma + \ln R_\rho \tag{3.2}$$

Applying (3.2) we obtain

$$\begin{aligned} S(\rho \mid \sigma) &= \operatorname{tr}(\rho(\ln \rho - \ln \sigma)) = -\operatorname{tr}\left(\rho^{1/2}(\ln \sigma)\rho^{1/2} - \rho^{1/2}(\ln \rho)\rho^{1/2}\right) \\ &= -\operatorname{tr}\left[\rho^{1/2}\left(\ln(L_\sigma)(\rho^{1/2}) + \ln(R_\rho)(\rho^{1/2})\right)\right] \\ &= -\operatorname{tr}\left[\rho^{1/2} \ln\left(\Delta(\sigma, \rho)\right)(\rho^{1/2})\right] \\ &= \left\langle -\ln\left(\Delta(\sigma, \rho)\right)(\rho^{1/2}), \rho^{1/2}\right\rangle \end{aligned} \tag{3.3}$$

In (3.3) we have written $S(\rho \mid \sigma)$ in terms of a single operator logarithm instead of two noncommuting operator logarithms and this is the key to our proof. We are now in position to prove the monotonicity inequality for quantum relative entropy. This inequality says that discarding a component of a compound quantum system can only decrease the relative entropy.

Theorem 3.2 *If $\rho_{12}, \sigma_{12} \in D_{nm}$ with $\rho_{12}, \sigma_{12} > 0$ are joint density operators and $\rho_1, \sigma_1 > 0$ are corresponding marginal states, then*

$$S(\rho_1 \mid \sigma_1) \leq S(\rho_{12} \mid \sigma_{12}) \tag{3.4}$$

Proof. Applying (3.3) we can rewrite (3.4) in the form

$$\left\langle -\ln\left(\Delta(\sigma_1, \rho_2)\right)(\rho_1^{1/2}), \rho_1^{1/2}\right\rangle \leq \left\langle -\ln\left(\Delta(\sigma_{12}, \rho_{12})\right)(\rho_{12}^{1/2}), \rho_{12}^{1/2}\right\rangle \tag{3.5}$$

We now define the linear transformation $U : M_n \to M_{nm}$ by

$$U(A) = (A\rho_1^{-1/2} \otimes I_m)\rho_{12}^{1/2}$$

We next show that $U^* : M_{nm} \to M_n$ is given by

$$U^*(B) = \operatorname{tr}_2\left[B\rho_{12}^{1/2}(\rho_1^{-1/2} \otimes I_m)\right] \tag{3.6}$$

To prove (3.6) we have that

$$\begin{aligned} \langle U(A), B\rangle &= \operatorname{tr}\left[(A\rho_1^{-1/2} \otimes I_m)\rho_{12}^{1/2}B^*\right] \\ &= \operatorname{tr}_1\left[A\operatorname{tr}_2\left((\rho_1^{-1/2} \otimes I_m)\rho_{12}^{1/2}B^*\right)\right] \\ &= \left\langle A, \operatorname{tr}_2\left[B\rho_{12}^{1/2}(\rho_1^{-1/2} \otimes I_m)\right]\right\rangle \end{aligned}$$

It is now demonstrated that U has the following properties:

(1) $U^*\Delta(\sigma_{12}, \rho_{12})U = \Delta(\sigma_1, \rho_1)$
(2) $U(\rho_1^{1/2}) = \rho_{12}^{1/2}$
(3) $U : M_n \to M_{nm}$ is an isometry.

To prove (1) we have that

$$\begin{aligned}\left[U^*\Delta(\sigma_{12}, \rho_{12})U\right](A) &= U^*\sigma_{12}(A\rho_1^{-1/2} \otimes I_m)\rho_{12}^{1/2}\rho_{12}^{-1} \\ &= \mathrm{tr}_2\left[\sigma_{12}(A\rho_1^{-1/2} \otimes I_m)\rho_{12}^{-1/2}\rho_{12}^{1/2}(\rho_1^{-1/2} \otimes I_m)\right] \\ &= \mathrm{tr}_2\left[\sigma_{12}(A\rho_1^{-1} \otimes I_m)\right] = \sigma_1 A\rho_1^{-1} \\ &= \Delta(\sigma_1, \rho_1)A\end{aligned}$$

To prove (2) we have that

$$U(\rho_1^{1/2}) = (\rho_1^{1/2}\rho_1^{-1/2} \otimes I_m)\rho_{12}^{1/2} = (I_n \otimes I_m)\rho_{12}^{1/2} = \rho_{12}^{1/2}$$

Finally, (3) can be proved as follows

$$\begin{aligned}U^*U(A) &= U^*\left[(A\rho_1^{-1/2} \otimes I_m)\rho_{12}^{1/2}\right] \\ &= \mathrm{tr}_2\left[(A\rho_1^{-1/2} \otimes I_m)\rho_{12}^{1/2}\rho_{12}^{1/2}(\rho_1^{-1/2} \otimes I_m)\right] \\ &= \mathrm{tr}_2\left[(A\rho_1^{-1/2} \otimes I_m)\rho_{12}(\rho_1^{-1/2} \otimes I_m)\right] \\ &= A\rho_1^{-1/2}\rho_1\rho_1^{-1/2} = A\end{aligned}$$

Hence, $U^*U = I_n$ so U is an isometry. We can now write (3.5) in the form

$$\begin{aligned}&\left\langle -\ln\left(U^*\Delta(\sigma_{12}, \rho_{12})U\right)(\rho_1^{1/2}), \rho_1^{1/2}\right\rangle \\ &\quad \le \left\langle -\ln\left(\Delta(\sigma_{12}, \rho_{12})\right)(\rho_{12}^{1/2}), \rho_{12}^{1/2}\right\rangle\end{aligned} \tag{3.7}$$

Applying Lemmas 2.1 and 2.2 we have that

$$-\ln\left(U^*\Delta(\sigma_{12}, \rho_{12})U\right) \le -U^*\ln\left(\Delta(\sigma_{12}, \rho_{12})\right)U$$

Hence,

$$\begin{aligned}
&\left\langle -\ln\left(U^*\Delta(\sigma_{12},\rho_{12})U\right)(\rho_1^{1/2}),\rho_1^{1/2}\right\rangle \\
&\quad\le\left\langle -U^*\ln\left(\Delta(\sigma_{12},\rho_{12})\right)U(\rho_1^{1/2}),\rho_1^{1/2}\right\rangle \\
&\quad=\left\langle -\ln\left(\Delta(\sigma_{12},\rho_{12})\right)U(\rho_1^{1/2}),U\rho_1^{1/2}\right\rangle \\
&\quad=\left\langle -\ln\left(\Delta(\sigma_{12},\rho_{12})\right)(\rho_{12}^{1/2}),\rho_{12}^{1/2}\right\rangle
\end{aligned}$$

which is (3.7). □

In Theorem 3.2 we assumed that all the density matrices were strictly positive. However, any density matrix can be approximated arbitrarily closely by a strictly positive density matrix. Since it is easy to show that $S(\rho \mid \sigma)$ is a continuous function of ρ and σ, we conclude that Theorem 3.2 holds for any $\rho_{12},\sigma_{12}\in D_{nm}$. Finally, we apply Theorem 3.2 to obtain an important inequality called *strong subadditivity* [2, 4, 6, 9].

Corollary 3.3 *If ρ_{123} is a joint density matrix for a composite of three quantum systems and $\rho_2,\rho_{12},\rho_{23}$ are corresponding marginal states, then*

$$S(\rho_{123})+S(\rho_2)\le S(\rho_{12})+S(\rho_{23}) \tag{3.8}$$

Proof. By the monotonicity inequality (3.4) we have that

$$S\left(\rho_{12}\mid\frac{I}{d}\otimes\rho_2\right)\le S\left(\rho_{123}\mid\frac{I}{d}\otimes\rho_{23}\right) \tag{3.9}$$

where I is the identity and d is the dimension of the first system. Writing (3.9) in terms of the definition of relative entropy and employing the appropriate partial traces gives

$$\begin{aligned}
-S(\rho_{12})+S(\rho_2)&=\operatorname{tr}(\rho_{12}\ln\rho_{12})-\operatorname{tr}(\rho_2\ln\rho_2)\\
&=\operatorname{tr}\left(\rho_{12}\ln\rho_{12}-\rho_{12}\ln\left(\frac{I}{d}\otimes\rho_1\right)\right)\\
&=S\left(\rho_{12}\mid\frac{I}{d}\otimes\rho_2\right)\le S\left(\rho_{123}\mid\frac{I}{d}\otimes\rho_{23}\right)\\
&=\operatorname{tr}\left(\rho_{123}\ln\rho_{123})-\rho_{123}\ln\left(\frac{I}{d}\otimes\rho_{23}\right)\right)\\
&=\operatorname{tr}(\rho_{123}\ln\rho_{123})-\operatorname{tr}(\rho_{23}\ln\rho_{23})\\
&=-S(\rho_{123})+S(\rho_{23})
\end{aligned}$$

Which is equivalent to (3.8). □

REFERENCES

[1] F. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley and Sons, New York, 1991.
[2] M. Nielsen and I. Chuang, *Quantum Computation and Quantum Information*, Cambridge University Press, Cambridge, 2000.
[3] E. Friedgut, Hyypergraphs, entropy, and inequalities, *Amer. Math. Monthly* **111** (2004), 749–760.
[4] M. Ohya and D. Petz, *Quantum Entropy and Its Use*, 2nd ed., Springer-Verlag, Berlin, 2004.
[5] A. Uhlmann, Relative entropy and the Wigner-Yanase-Dyson-Lieb concavity in an interpolation theory, *Commun. Math. Phys.* **54** (1977), 21–32.
[6] E. H. Lieb and M. B. Ruskai, Proof of the strong subadditivity of quantum mechanical entropy, *J. Math. Phys.* **14** (1973), 1938–1941.
[7] S. Gudder, Quantum computation, *Amer. Math. Monthly* **10** (2003), 181–201.
[8] R. Bhatia, Matrix Analysis, *Springer-Verlag* , New York, 1997.
[9] M. Nielsen and D. Petz, A simple proof of the strong subadditivity inequality, arxiv: quant-ph/0408130 (2004).
[10] D. Petz, Quasi-entropies for finite quantum systems, *Rep. Math. Phys.* **23** (1986), 57–65.

RICHARD HEALEY*

7. SYMMETRY AND THE SCOPE OF SCIENTIFIC REALISM

ABSTRACT

Scientific realism characteristically encompasses semantic as well as epistemological claims. When a theory of modern physics is empirically successful, the realist considers this a reason to believe that it is approximately true, and that terms of the theory (including those it newly introduces) typically refer to or represent real physical structures—irrespective of whether these are accessible to our senses or independently measurable. But suppose every model of a theory may be mapped into a distinct model by a transformation that preserves all measurable structures. Then the empirical success of the theory fails to support realist claims about purportedly distinct structures related by such a symmetry. Assuming there are such structures, the theory's success provides no reason to believe that terms it introduces determinately refer to or represent them, and no basis for any specific belief about them. Even a scientific realist then has no grounds for thinking there are any such structures. These include Newtonian absolute space and the gauge potentials of classical electromagnetism acting on classical or quantum charged particles.

1 INTRODUCTION

Suppose a scientific theory T is newly proposed for some domain D, in part by exhibiting a class of mathematical structures ("models") regarded as appropriate for representing what happens in D. Prior to this proposal, D was described using a language L_O. It may be that some elements of models of T purport to represent physical structures that are referred to by no term of L_O. A term referring to such a structure may still be explicitly definable in L_O. But typically no definition will be to hand when models of T represent physical structures T newly posits as "lying behind" the phenomena in D. Examples are familiar from fundamental physics in which new terms are introduced into the language, or existing terms are deliberately used in new ways ('quark', 'Higgs field', 'color charge'), precisely in order to refer to such novel structures.

This situation raises two worries, one semantic, the other epistemological. How does an undefined term, newly introduced with a theory, get its meaning? Why believe that there is anything in the world to which such a term refers?

* Richard Healey (2005). Dept. of Philosophy, University of Arizona, Tucson AZ 85721, USA; Email: rhealey@email.Arizona.edu

W. Demopoulos and I. Pitowsky (eds.), Physical Theory and its Interpretation, 143–160.

Holism attempts to address these worries. According to semantic holism, the primary vehicle of meaning is not the individual theoretical term, nor even statements containing such terms, but the entire theory that introduces them. As long as the theory itself is meaningful, there can be no residual doubt as to the meaningfulness of its constituent terms: their meaning accrues to them by virtue of their role in the theory. The theory derives its meaning from its empirical content—from the testable claims it makes about its domain. But according to confirmational holism, it is the whole theory, rather than individual theoretical claims, that "faces the tribunal of sense experience" (in Quine's (1951, [1953, p. 41]) famous phrase). If the theory is favorably judged by this tribunal, then each of its theoretical claims is worthy of belief, for none is testable in isolation from the rest of the theory. Given the success of quantum chromodynamics, we should accept that theory, and believe there are quarks with the features it attributes to them, including their color charges. 'Quark' simply refers to quarks, and 'color charge' to their color charges. According to semantic holism, one comes to understand what such terms mean just by learning the theory of quantum chromodynamics in which they figure, including the successful applications that warrant its acceptance. No definitions in pre-existing language are needed.

But holism fails to quiet the worries. It is now widely acknowledged that the models of a theory may contain surplus structure (Redhead (1975))—structure that may be eliminated without loss to the theory's empirical content or super-empirical virtues.

It is easy to construct fanciful examples by adding elements to the models of an existing theory—elements that add nothing to the theory's empirical content while purporting to represent structures that would be epistemically inaccessible even if they existed. For instance, one can add a scalar field to models of Maxwell's electromagnetism that has no interactions with particles (charged or neutral) or other fields, and carries no energy or momentum. No measurement could probe the values of any such field, and its presence would have no empirical consequences.

Real examples are more interesting if they can be found. But it is often difficult to determine that theoretical models contain surplus structure. Moreover, such determinations remain hostage to further scientific developments that may connect such "idly turning wheels" to other theoretical mechanisms, rendering what they represent epistemically accessible by providing theory-mediated links to observation. Two theories whose models have often been thought to contain surplus structures are Newtonian mechanics (with its absolute space) and electromagnetism (in a formulation that takes potentials rather than fields as basic). Each of these theories has distinctive features that merit the individual treatment offered in a later section of this paper (§6, §7 respectively). But each is also arguably an instance of a general class of theories in which surplus structure is identifiable by appeal to a symmetry of a theory—a mapping that takes one model of the theory into another that differs from it only with respect to the structure in question. Ismael and van Fraassen (2003) argue that surplus structure can often be detected in a theory precisely by identifying such symmetries.

Whatever its merits, holism does not provide the necessary tools for an adequate analysis of the epistemological role of surplus structures, or of the semantics of terms purporting to refer to what these structures represent. If present in theoretical models,

surplus structure cries out for special treatment precisely because its lack of integration with other model structures naturally breaks up the content of a theory so that it is no longer adequate to treat it as an undivided whole. If a term newly introduced with a theory purports to refer to what is represented by surplus structure in that theory's models, then holism cannot quiet our two worries. How, if at all, does such a term get its meaning? And can the success of the theory provide any reason to believe that it has a reference?

An influential paper by David Lewis (1970) may seem to contain an answer to the first, semantic, question. In that paper Lewis outlines a general procedure for generating definitions in a pre-existing language L_O for terms newly introduced by a theory T. After a brief review of the application of this procedure to the present issue in section 2, section 3 explains why Lewisian definitions fail to specify determinate meanings for newly introduced terms in the presence of just the kinds of theoretical symmetries that are the mark of surplus structure. Section 4 generalizes a subsequent suggestion of Lewis that demonstration may secure determinate meanings in such a case. But for certain theories even demonstration may fail to remove residual indeterminacies. Section 5 argues that this failure is not a vice but a virtue of Lewis's procedure. Its application here exhibits the semantic indeterminacy of newly introduced theoretical terms purporting to refer to anything such surplus structures represent. This semantic indeterminacy is a symptom of an epistemological defect in the introducing theory. Even if there were something in the world corresponding to surplus structure in its models, a proponent of the theory could neither say just what it was nor form determinate beliefs about it. As long as this situation persists, no amount of empirical success of the theory could warrant belief that there is anything represented by such surplus structure. These general lessons are then applied to specific theories. Section 6 argues that there is an important sense in which Newton did not know what he was talking about when discussing absolute space. Section 7 shows why a realist who accepts classical electromagnetism would still have no reason to believe that any localized properties are represented by electromagnetic potentials, and no way even to entertain suggestions as to just which properties are localized where.

2 HOW LEWIS PROPOSED TO DEFINE THEORETICAL TERMS

Lewis (1970) outlined a procedure for using the Ramsey sentence of a theory T to construct explicit definitions for terms newly introduced by T in a previously understood "O-language" L_O. It may seem surprising that anything like this is possible, given the perceived failure of attempts by the logical positivists to define theoretical terms in (what they called) an observation language, and work of Suppes (1957) and others establishing (against claims of Mach) the indefinability of a term like 'mass' within classical particle mechanics. The success of Lewis's method depends not only on a liberal understanding of the logical and linguistic resources available to provide the required definitions, but also on a certain substantive assumption that the theory be *uniquely realized*. It is essentially this assumption that is called in question when symmetries of T manifest surplus structure in its models. If it fails, then so does

Lewis's method for defining theoretical terms. To understand why, it is necessary to explain Lewis's procedure and to say just what the assumption of unique realization amounts to in this context.

Lewis begins by assuming that, given a successful new theory T, it is possible to formulate it by means of a single (finite or infinite) postulate of the form '$T[\tau_1 \ldots \tau_n]$', where $\tau_1 \ldots \tau_n$ are all the terms newly introduced by T, while all the other terms appearing in the postulate are assumed to be already understood, with determinate sense and reference. In defense of this assumption, he argues that if new terms appear in the theory as predicates, functors, etc. one can replace them in these occurrences by names of corresponding properties, functions, etc. by making use of already understood copulas. For example, a new monadic predicate '$\Phi\_$' could be replaced by a term '$\_$ has the property ϕ' where 'ϕ' purports to name the property (necessarily) shared by all and only objects of which 'Φ' is true, and '$\_$ has $\_$' is a previously understood relational predicate. He calls the result '$T[x_1 \ldots x_n]$' of replacing the terms $\tau_1 \ldots \tau_n$ by distinct variables $x_1 \ldots x_n$ that do not already occur in T the *realization formula* for T, so that any n-tuple of entities that satisfies it (keeping the interpretation of all other terms in T fixed) *realizes* or is a *realization* of T. So the postulate T says that T is realized by the n-tuple of entities denoted (respectively) by $\tau_1 \ldots \tau_n$. The *Ramsey sentence* R, i.e. '$\exists x_1 \ldots \exists x_n T[x_1 \ldots x_n]$' , on the other hand, merely says that T is realized by at least one n-tuple of entities. But any consequence of T in the language L_O available before the introduction of the new terms $\tau_1 \ldots \tau_n$ by T is still a consequence of R, and so the logically weaker R shares all the same predictions, and therefore predictive success, as T, while containing none of its new terms. But of course, R is still committed to the existence of entities corresponding to these new terms, since it will not be true unless some n-tuple of entities realizes T.

If a theory is uniquely realized, then the following identities will serve to define its newly introduced terms in previously understood language,

$$\tau_1 = \iota y_1 \exists y_2 \ldots \exists y_n \forall x_1 \ldots \forall x_n (T[x_1 \ldots x_n] \equiv (y_1 = x_1) \& \ldots \& (y_n = x_n))$$

.....

$$\tau_n = \iota y_n \exists y_1 \ldots \exists y_{n-1} \forall x_1 \ldots \forall x_n (T[x_1 \ldots x_n] \equiv (y_1 = x_1) \& \ldots \& (y_n = x_n))$$

where the symbol 'ι' is the description operator: 'ιx' stands for 'the object x such that'. The identities fix the references of the newly introduced terms by saying what they actually denote: and they specify their senses by saying what they would denote if T were uniquely realized by some other n-tuple of entities. Lewis (1970,433) says this about the assumption of unique realization.

> A uniquely realized theory is, other things being equal, certainly more satisfactory than a multiply realized theory. We should insist on unique realization as a standard of correctness unless it is a standard too high to be met. Is there any reason to think that we must settle for multiply realized theories? I know of nothing in the way scientists propose theories which suggests that they do not hope for unique realization. And I know of no

> good reason why they should not hope for unique realization. Therefore I contend that we ought to say that the theoretical terms of a multiply realized theories (sic) are denotationless.

Lewis's proposal applies directly only to theories formulated as a single theoretical postulate. But a physical theory is rarely, if ever, formulated in this way—as a sentence in some definite language that can be regarded as a (possibly infinite) conjunction of its laws or basic principles. Indeed a historian seeking any definitive statement of a physical theory is rapidly struck by the diversity of languages, mathematical structures and basic principles that scientists have used in formulating what they consider the same theory. A philosopher wishing to understand the structure and function of a physical theory must idealize.

A degree of consensus has emerged in contemporary philosophy of science that an illuminating idealization is to be found in some version of the semantic conception of scientific theories. On this conception, a physical theory may be idealized as presenting a collection of *models* intended to represent structures in its intended domain of application. A model in this sense can be thought of as a mathematical structure satisfying certain conditions: it will be convenient to take it to have the canonical form of an n-tuple $< D, Q_1, Q_2, \ldots, Q_n >$, where D is a domain of objects (abstract, concrete, or both) and each Q_i is a magnitude, i.e. a function from some subset of D^m into the real numbers, or some other mathematical space. For example, models of the general theory of relativity are often given in the form of triples $< M, g, T >$, where M is a differentiable manifold, g is a metric tensor field on M, and T is a stress energy tensor field. This does not have the canonical form as it stands, but it could be brought into that form with a little work.

While a collection of such models makes no assertions by itself, it does provide resources for a scientist to make theoretical claims. Just what form these should take, and what is the appropriate epistemic attitude to adopt toward them, are still to be decided. The general idea is to claim that the world, or some part(s) or aspect(s) of it, has(have) more or less the same structure as some model(s), or part(s) thereof; and to propose adoption of a suitably favorable epistemic attitude to such claims just in case they are well enough supported by observation and experiment. Of course, this vagueness covers a host of contentious issues in the philosophy of science that cannot and need not be resolved here.

For present purposes it is necessary only to note that for the scientific realist, the claim of similarity of structure of model to world extends beyond similarity to those features that are observable or measurable independently of the theory. The realist maintains that observation and experiment may warrant further belief, to the effect that entities and/or magnitudes newly introduced into some of the theory's models represent physical structures that were neither known nor knowable prior to formulation of the theory. If this belief is true, then we should expect to be able to entertain specific beliefs and to formulate testable claims about these structures. The expectation can be met only if there is some way of determinately referring to them. This is how Lewis's concerns arise within the semantic conception of scientific theories.

Consider the claim that some aspect of reality has the same structure as a model $m =< D, Q_1, Q_2, \ldots, Q_r >$ of a physical theory. For example, one might claim that Mercury's orbit around the sun has the same structure as the trajectory of a massive particle in the Schwarzschild solution to the field equations of general relativity. Or one might make the completely general claim that the theory of general relativity is true, in the sense that the universe has the structure of some model of general relativity. Any such claim C may be made precise by formulating it as the assertion of an isomorphism of a certain kind between a model m and a real-world structure $< W_1, W_2, \ldots, W_r >$. The claim C is the analog of Lewis's postulate T, and (after reordering) the terms $W_1, W_2, \ldots, W_n (n \leq r)$ purporting to name hitherto unknown physical structures are the analogs of Lewis's terms $\tau_1 \ldots \tau_n$. We can now formulate the Ramsey sentence of $C[W_1, W_2, \ldots, W_n]$ as the claim '$\exists x_1 \ldots \exists x_n C[x_1 \ldots x_n]$'. C is uniquely realized just in case a unique sequence of physical structures satisfies this Ramsey sentence, in which case $W_1, W_2, \ldots, W_n$ succeeds in naming this unique sequence. Given a collection of models associated with a theory, one can now ask whether a particular theoretical claim based on these models is uniquely realized. If it is, then Lewisian definitions of the associated theoretical terms $W_1, W_2, \ldots, W_n$ will be forthcoming. Note however, that unless the theoretical claim is completely general these can only partially specify their extensions and intensions.

3 SYMMETRY AND MULTIPLE REALIZATION

Lewis's uniqueness assumption fails for an important class of theories with the right kinds of symmetries. It is therefore worth noting that after his (1970) Lewis changed his attitude toward this assumption. For example, in his (1994, [1999,301]) he considers the possibility of multiple realizations of a folk-psychological theory of mind containing a term M. He says

> I used to think that in this case the name M had no referent. But now I think it might be better, sometimes or always, to say that the name turns out to be ambiguous in reference. That follows the lead of Field (1973)

Field (1973) considered a case in which there are no grounds for concluding that a theoretical term in a superseded theory (say Newton's term 'mass') denotes one thing rather than another denoted by distinct terms of a replacing theory (say Einsteinian 'rest mass' and 'relativistic mass'). He argued that one should then say that the offending term partially denotes each of the rival candidates: and that theoretical sentences containing the term should count as true if and only if they came out true under each partial denotation (essentially following van Fraassen's (1966) method of supervaluations). In his (1994, 1997) Lewis contemplates the possibility of multiple realizations of a theory only in cases in which some 'deeper' theory is available to describe such alternative partial denotations.

In a more recent (posthumous) paper, he again deploys the same basic mechanism for defining theoretical terms for a different purpose, namely to argue that

> Quite generally, to the extent that we know of the properties of things only as role-occupants, we have not yet identified those properties. No amount of knowledge about what roles are occupied will tell us which properties occupy which roles.

This time the mechanism is applied not just to a particular scientific theory, but rather to a "true and complete 'final theory'" capable of delivering "a true and complete inventory of those fundamental properties that play an active role in the actual workings of nature". My concern is limited to particular scientific theories, which are presumably neither true nor complete. But there is still a certain commonality of argumentative strategy here, since I wish to argue that a proponent of a theory in which symmetries are a mark of surplus structure is caught in the very similar predicament of being unable to say which such structures occupy which roles in any actual situation.

In contrast to his earlier papers, Lewis (1999) now thinks that he can secure unique realization for his true and complete "final theory" by a simple move. His new thesis is that there is no way to distinguish this from its multiple *possible* realizations.

> Though our theory T has a unique actual realization, I shall argue shortly that it has multiple possible realizations. Suppose it does indeed have multiple possible realizations, but only one of them is the actual realization. Then no possible observations can tell us which one is actual, because whichever one is actual, the Ramsey sentence will be true. There is indeed a true contingent proposition about which of the possible realizations is actual, but we can never gain evidence for this proposition, and so can never know it.

The new thesis and Lewis's argument for it are not my concern, which is the possibility of multiple actual realizations of a theory whose models contain surplus structure. But this makes it important to consider Lewis's (1999) new reason for dismissing such a possibility. Here is what he says

> We have assumed that a true and complete final theory implicitly defines its theoretical terms. That means it must have a unique actual realization. Should we worry about symmetries, for instance the symmetry between positive and negative charge? No: even if positive and negative charge were exactly alike in their nomological roles, it would still be true that negative charge is found in the outlying parts of atoms hereabouts, and positive charge is found in the central parts. O-language has the resources to say so, and we may assume that the postulate mentions whatever it takes to break such symmetries. Thus the theoretical roles of positive and negative charge are not purely nomological roles; they are locational roles as well.

The idea seems to be to secure unique realization for the terms 'positively charged' and 'negatively charged' in face of the assumed symmetry of the fundamental theory

in which they figure by adding one or more sentences stating what might be thought of as "initial conditions" to the laws of that theory. These sentences S would be formulated almost exclusively in what Lewis calls the O-language—i.e. the language L_O that is available to us without benefit of the term-introducing theory T. But they would also use one or more of the terms 'positively charged' and 'negatively charged' to break the symmetry of how these terms figure in T. They would do this by applying further constraints that must be met by the denotations of these terms in order that $S\&T$ be true. Those constraints would then fix the actual denotations of 'positively charged' and 'negatively charged' in T so that, subject to these further constraints, T is indeed uniquely realized.

Lewis expresses an important insight here. While a fundamental theory in physics is concerned to capture universal laws governing the workings of the world, to apply this theory to a particular situation it must be possible to use the theory to describe or represent that situation. If this were not possible, the theory would be useless. Moreover, we could have no reason to believe it, since observations of particular situations could provide no evidence for the theory. Applications of the theory provide the resources to set further constraints on the denotations of its newly introduced terms—constraints that may suffice to break the symmetries of its laws and so secure its unique realization. Note that such constraints need not involve descriptions in L_O, though they typically will. But they will involve demonstration or ostension, as does Lewis's own suggestion when it includes the term 'hereabouts'. One could, for example, simply point to a cathode and say "That is negatively charged".

Essentially the same point may be made within the semantic conception of scientific theories, in which the theory T is true just in case the physical world has the structure of one of a specified collection of models. For it may be that T is symmetric under exchange of positive with negative charge, i.e. that the object mN that results from systematically interchanging magnitudes of positive and negative charges in any model m in this collection is also in the collection. Assume that some claim is warranted to the effect that an actual physical structure s is isomorphic to a particular model m *via* the map i ($s = i(m)$ for short), and that the image of m under a systematic interchange of positive with negative charge is $m' = h(m) \neq m$. Then m is also isomorphic to s *via* the map $i' = i \circ h$, and so the claim is multiply realized. But now we can *stipulate* that the magnitude in m that i rather than i' maps into a uniquely individuated physical quantity of one or more objects in s has *negative* charge. Again, the symmetry is broken by demonstration of a particular physical structure. And again this permits one to define newly introduced terms such as positive and negative charge in such a way as to secure determinate denotations for these terms.

4 DEMONSTRATION AND ITS LIMITS

There is an important general lesson to be drawn from Lewis's example of electric charge in a theory with charge symmetry. The available resources for specifying the denotations of newly introduced terms are not purely descriptive. Even when the purely descriptive claims of a new theory are multiply realized, it may be possible to

fix on a unique realization of that theory by demonstration.[1] It may: but then again, it may not. Demonstration has its limits.

Consider, for example, the following toy theory.[2] Suppose that physicists in a possible world not too different from our own try to account for the properties of the strong nuclear force in their world. They arrive at a classical (not quantum) theory modeled on classical electrodynamics, but resembling chromodynamics in that it postulates three different "color" charges (along with their opposites). The physicists postulate that the building blocks of matter are quarks, each of which bears a smallest unit of color charge. They formulate detailed dynamical laws governing the behavior of particles under the strong force. These laws are completely symmetric under permutations of color charge, and also imply that quarks will always be confined within color-neutral combinations. Nucleons are taken to consist of three confined quarks, each of a different "color", while mesons are composed of an oppositely colored quark and antiquark pair. Confinement is very strong in this theory. For example, the three quarks in a nucleon are point particles that always occupy exactly the same point of space as each other. The theory can model the dynamics of free quarks, including how appropriate combinations would "collapse" into color-neutral point combinations. But in fact there have never been any free quarks; and, because of strong confinement, there never will be.

This theory could be applied to explain detailed properties of nuclei in this world, as well as predicting cross-sections for various scattering processes, such as the production of pi-mesons in proton-proton collisions. It could prove very successful in such applications, and could come to be believed on the basis of that success. But because of its color symmetry, the theory would not have a unique realization in that world. Moreover, because of permanent, strong confinement, there would be no way in that world to say or demonstrate which quarks in a nucleon are "green", which "blue" and which "red", even though (for example) every nucleon was known to consist of precisely one quark of each color. So Lewis's move would be to no avail: multiple realization would be unavoidable.

In this example, demonstration cannot constrain realizations of a theory that is symmetric under permutations of color charge. But the example is a little delicate. It depends upon the fact that, in this case, demonstration is *itself* indeterminate. Add just two temporarily isolated and differently color-charged quarks to the world (or even two mesons, known to be composed of quark antiquark pairs of different color-charges), and demonstration may secure a unique realization. It will do so as long as *any* realization fixes the (same) denotation for the relation *has the same color charge as*. This provides an illuminating contrast with the case of classical electromagnetism to be considered later (in section 7).

But feasible demonstrations can eliminate some candidates for realizing a newly introduced theory with a certain symmetry, while still failing to secure a unique realization. Whether feasible or not, the number of independent demonstrations that would be required to single out just one realization will depend on the structure of the theory. It is here that an important connection to measurability manifests itself. When part of the denotation of a newly introduced theoretical term is fixed by a

demonstration, there can be no question of a measurement revealing that the term does not apply to the demonstrated item. In Lewis's example, for a theory that is symmetric under exchange of newly introduced terms referring to positive and negative charges, the demonstrative act "That cathode is negatively charged" functions as a stipulative definition. Nothing could count as a measurement as to whether that cathode is positively or negatively charged while the theory and the stipulation remain sacrosanct.

Measurements of whether *other* objects have positive or negative charge is possible, given the theory. For example, if the theory is true, then bringing a similarly charged object near the cathode will have detectable effects that are different from those that ensue when an oppositely charged object is brought near. For certain objects, those effects may be directly observable. Detection of differential effects in the case of other objects will be more indirect, and may rely not only on observation, but also on other theories that were in place before the charge-introducing theory was proposed. But whether direct or indirect, the possibility of measuring whether the charge of a second object is positive or negative depends on two assumptions. The first assumption is that the charge of *that* object was not also fixed by stipulation. The second assumption is that the denotation of the term 'negatively charged', secured (in part) by stipulating that a particular cathode was negatively charged, extended to objects distinct from that cathode, including this second object. It is the term-introducing theory that effects this extension, if anything does, by determining the denotation of 'similarly charged' by a Lewisian definition. And because this determination is given by a uniquely satisfied Ramsey sentence, a measurement of whether the second object is positively or negatively charged is possible: one just has to determine whether or not the conditions specified in L_o are met by the chosen cathode and the second object.

In the toy color-charge theory, a stipulation to the effect that a single isolated quark (or the quark in a quark-antiquark pair composing a pi meson) is red would not suffice to secure a unique realization in a world containing non-red quarks. This stipulation would fix the denotation of 'red quark' and thereby ground contingent claims about whether other demonstrable quarks are red. But without further stipulation the denotations of 'blue quark' and 'green quark' would remain indeterminate. No measurement could reveal whether a particular demonstrated quark was blue rather than green—not because of the epistemic inaccessibility of these color charges, but because no assertion or belief about its result has been given the required determinate content. If you can't say it, you can't measure it either.

5 MULTIPLE REALIZATION AS A GUIDE TO SURPLUS THEORETICAL STRUCTURE

The examples of theories introducing electric or color charge showed how it is possible for a theoretical claim or postulate to be multiply realized because of symmetries that map a model representing one realization into a distinct model representing a different realization while preserving all measurable magnitudes. In such a case, Lewis's

original (1970) proposal will fail to assign a determinate denotation to some theoretical term(s) introduced with that theory. Section 4 generalized Lewis's (forthcoming) subsequent suggestion to describe how demonstrative stipulation may render their denotations more determinate. In some cases this can help to answer the semantic question as to how these terms get their meanings. But demonstration can succeed only if the theory is in fact realized; and even when demonstration is successful, the need for it points to a residual epistemological problem for the theory. The problem arises whenever multiple realization consequent upon theoretical symmetries threatens semantic indeterminacy, whether or not that threat may be averted by stipulative demonstration. In this way Lewis's (1970) proposal becomes a diagnostic tool for locating epistemological defects in a theory.

Consider Lewis's example of a charge-symmetric theory T that newly introduces the terms 'positively charged' and 'negatively charged'. Assume that the Ramsey sentence of T is true, and that at least one n-tuple of entities realizes T, some of which are charged. A global switching of positive with negative charge results in a distinct n-tuple of entities that realizes T. So T is multiply realized. We can settle on a unique realization by making a stipulative demonstration to the effect that *this* or *these* entities are negatively charged, and this will resolve any semantic indeterminacy in the terms 'positively charged' and 'negatively charged'. But the need to do so highlights the fact that T contains surplus structure. T has no need to postulate two distinct intrinsic properties corresponding to positive and negative charge. Given its charge symmetry, there is a "leaner" theory T' that postulates only an external relation 'being oppositely charged to'[3]. Unlike T, T' will be uniquely realized. The stipulative demonstration that served within T to fix certain intrinsic properties as the denotations of the terms 'positively charged' and 'negatively charged' serves within T' merely to introduce convenient labels for classes of oppositely charged entities. T' itself has no need of such terms, for it postulates no intrinsic properties for them to denote.

The example generalizes. Any theory with symmetries that preserve all measurable structures will be multiply realized. Stipulative demonstrations may eliminate some realizations, or even all but one, thereby reducing the semantic indeterminateness of newly introduced theoretical terms. But this need for them is a symptom of an epistemological defect in the theory. Its models contain surplus structure—elements purporting to represent real structures but that play no role in contributing to the theory's success. It may be difficult or practically impossible to eliminate such structures, and it may be convenient to retain them as an aid to calculation or as a fruitful heuristic guide to new theory construction. But as long as the theory retains the symmetries in question, the continued success of the theory will provide no reason to believe that there is anything in the domain of the theory for these structures to represent.

There is an objection that must be addressed at this point. The objection is that Lewis's (1970) proposal cannot constitute a complete account of the meanings of newly introduced theoretical terms. For if no constraints are placed on the denotations of newly introduced theoretical terms other than that T be consistent, then Twill always be trivially multiply realized in any domain of the right cardinality.[4]

But T was a scientific theory, capable of being false if T is not realized. Therefore there must be additional constraints on its newly introduced terms that Lewis's (1970) proposal fails to reveal.

There are at least two responses to this objection. Lewis (1984, [1999,65–68]) attempts to address the acknowledged incompleteness of his account of the meanings of newly introduced theoretical terms by restricting their denotations to "natural" classes or properties. While Cruse (2005) points out that the consistency of T is not the only relevant constraint if the denotation of "old" terms is taken to be fixed even on a domain of "unobservable" entities that T purports to describe.

But whatever its merits in general, this is not a good objection to the present use of Lewis's (1970) proposal. To use that proposal as a diagnostic tool for probing for epistemological weaknesses in a theory, it is not necessary to defend Lewis's (1984) or any other any account of the nature and origin of whatever further constraints render non-trivial the claim that a theory introducing new terms is realized. All that matters is that *whatever* these constraints may be, and however many trivial unintended realizations they exclude, they still fail to rule out the multiple realizations of a theory consequent upon symmetries that preserve all measurable structures.

6 NEWTON'S ABSOLUTE SPACE

Newton set his theory of mechanics and gravitation within a framework of an unobservable, 3-dimensional Euclidean space that endured through a 1-dimensional Euclidean time and in which bodies were located. This framework grounded the basic concepts of straight line, distance, rest, uniform velocity, and acceleration apparently required by his theory. As Newton was the first to admit, his enduring absolute space presented epistemological difficulties.

> It is indeed a matter of great difficulty to discover and effectually to distinguish the true motions of particular bodies from the apparent; because the parts of that immovable space in which those motions are performed do by no means come under the observation of our senses. Yet the thing is not altogether desperate[5]

It has long been recognized that a framework with less structure suffices to formulate Newton's theory,[6] and that this removes at least some of the epistemological difficulties. The framework incorporates no distinguished state of rest, among all states of uniform motion. That such a rest state constitutes surplus structure in Newton's theory follows from the symmetries of the theory—mappings that preserve all measurable structures while varying the state of rest. The theory is therefore multiply realized, if true, giving rise to just the expected kinds of semantic indeterminacy. While these may be removed by stipulative demonstration, this would likely not have been acceptable to Newton himself. It follows that, even if his theory had been true, there is an important sense in which Newton would not have known what he was talking about when referring to an enduring absolute space.

Following Friedman (1983, 113), I take a model of Newtonian kinematics to be a sextuple $< M, D, dt, h^{ab}, V^a, T_\sigma >$, where M is a 4-dimensional differentiable manifold, and D is a flat affine connection on M that is compatible with the co-vector field dt (defining intervals of absolute time elapsed between point-events represented by points of M), the symmetric tensor field h^{ab} (defining a Euclidean metric on space at each instant) and the vector field V^a that defines the state of rest by specifying which point of M represents any given point of absolute space at each instant: T_σ are the tangent vectors to the curves σ in M representing where each particle is at each instant. If $< M, D, dt, h^{ab}, V^a, T_\sigma >$ is a model, then so is $< M, D, dt, h^{ab}, V'^a, T_\sigma >$, where $V'^a = f(V^a)$, and f is a Galilean transformation (a combination of a velocity boost, a spatio-temporal translation, and a spatial rotation). If the velocity boost is non-zero, then Newton would take these models to represent particles in different states of true (but not apparent) motion: if a particle is at rest according to the first, then it is moving with constant velocity according to the second, and *vice versa*.

But according to the theory these states are indistinguishable—there is no magnitude whose measurement could discriminate between them. If one is dynamically possible, then so is the other (assuming that all forces are, like Newtonian gravity, invariant under Galilean transformations). They have the same geometry and chronometry. And they agree on the relative motions of all particles. Only a direct determination of the state of rest could distinguish them, but that is impossible since points of space are not themselves observable. It follows that if Newton's theory is true, then it is multiply realized. The attempt to use Lewis's (1970) procedure to define the term *is at rest* will leave its denotation indeterminate, no matter what particles and forces there are. This pinpoints the epistemic deficiency of the theory: the structure V^a in its models is surplus, and should not be taken to represent anything real to which that term corresponds, no matter how successful the theory may be. It can and should be simply omitted from the models, resulting in a 'trimmed down' theory that no longer postulates an enduring 3-dimensional space, even though it still incorporates a distinction between accelerated motion and motion with constant velocity.

Within this "trimmed down" theory one may choose to privilege some particular state of uniform motion and *call* it a state of rest by a stipulative demonstration. But this adds no additional content to the theory: it is not a contingent claim that some particular unaccelerated object is at rest.

This is not apparently how Newton understood his own theory. In Book III of the *Principia* he introduces what he calls a hypothesis "acknowledged by all"—that the center of the system of the world is immovable, and quickly qualifies this consistent with his theory to mean that the center of gravity of the solar system is immovable. But if this is indeed intended as a contingent hypothesis, then it cannot be functioning as a stipulative demonstration of the state of rest. Newton seems to have assumed that the term *is at rest* has a determinate denotation independently of any such stipulative definition. The assumption seems plausible if one restricts attention to cases in which it corresponds to a *relation* between bodies or other observable items. Assuming they are uniformly moving, these items are relatively at rest just in case

they maintain a constant distance from one another. But the plausibility evaporates if one of the *relata* is a point of space. Even if there are enduring points of space, they are admittedly unobservable, and Newton's theory endows them with no causal powers. I conclude that Newton could not determinately have referred to an enduring, immovable, absolute space even if his theory had been true. For, given that theory, there can be no descriptive, causal or any other mechanism by which such determinate reference could be secured. Newton's hypothesis—that the center of gravity of the solar system is immovable—along with many of his other assertions and beliefs about absolute rest and absolute space would have had no determinate content even if his theory had been true.[7]

7 ELECTROMAGNETIC POTENTIALS

Classical electromagnetic theory presents another example of surplus structure when formulated in terms of potentials. Here the multiple realization consequent upon gauge symmetry is even more extreme. This leads to radical indeterminateness in denotation even assuming the truth of the theory. Moreover, such indeterminateness cannot be rectified by stipulative demonstration without emptying specific assertions or beliefs about the values of magnitudes of empirical content.

A model of classical electromagnetic theory applicable to phenomena in a vacuum specifies the values of two vector fields, $\boldsymbol{E}$ and $\boldsymbol{B}$, representing electric and magnetic field strengths respectively, together with a scalar field ρ representing charge density, and a vector field $\boldsymbol{j}$ representing current density, at each point in Minkowski space-time. The values of these magnitudes are required to satisfy Maxwell's equations. The relativistic covariance of the theory is made manifest by defining $\rho, \boldsymbol{j}$ to be the components of a Lorentz 4-vector j^μ, and $\boldsymbol{E}$ and $\boldsymbol{B}$ to be components of an antisymmetric electromagnetic field tensor $F^{\mu\nu}$. In a formulation that takes electromagnetic potentials rather than fields as basic magnitudes, the electromagnetic field strength is *defined* in terms of a 4-vector electromagnetic potential A^μ by the equation

$$F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu \tag{7.1}$$

It is invariant under the gauge transformation

$$A'^\mu(x) = A^\mu(x) + \partial^\mu \Lambda(x) \tag{7.2}$$

where $\Lambda(\boldsymbol{x}, t)$ is an arbitrary, but suitably differentiable, scalar function.

Classically, the effects of electromagnetism become manifest only through the action of electromagnetic fields on charged particles. A model of the theory specifies their trajectories by means of time-like curves in the manifold representing space-time: their velocities $\boldsymbol{v}$ are required to satisfy the Lorentz force law.

$$m\frac{d\boldsymbol{v}}{d\mathrm{t}} = q(\boldsymbol{E} + \boldsymbol{v} \times \boldsymbol{B}) \tag{7.3}$$

Since this involves the 4-vector potential A^μ only through the gauge invariant field strength $F^{\mu\nu}$, the trajectories of charged particles are invariant under gauge transformations. If $< M, \eta, A^\mu, J^\mu, T_\sigma >$ is a model of classical electromagnetism with sources J^μ producing an electromagnetic potential that acts on charged particles of charge q with trajectories T_σ in otherwise empty space-time $< M, \eta >$ with Minkowski metric η, then so is $< M, \eta, A'^\mu, J^\mu, T_\sigma >$, where A^μ and A'^μ are related by an arbitrary gauge transformation of the form (7.2).

Assume that some model of this theory exactly matches all trajectories of charged particles. Then so does any of a continuous infinity of other models related by (7.2). This gauge symmetry of the theory preserves all measurable magnitudes. It is therefore multiply realized. This a symptom of surplus theoretical structure. Understood realistically, the theory is epistemologically defective, because it postulates a theoretical structure that is not measurable even if the theory is true. The gauge-related models $< M, \eta, A^\mu, J^\mu, T_\sigma >$ and $< M, \eta, A'^\mu, J^\mu, T_\sigma >$, should therefore be regarded as representing the same physical situation, which may be uniquely represented by the model $< M, \eta, F^{\mu\nu}, J^\mu, T_\sigma >$ of a "stripped down" theory in which only the fields $F^{\mu\nu}$ are taken to represent genuine physical magnitudes, while any compatible 4-vector potential A^μ that is introduced is regarded as simply a mathematical convenience, with no further representative role. Moving to the "stripped down" theory removes the epistemological defect, and gauge symmetry no longer presents a threat to unique realization of the new theory.

Nevertheless, the original theory could have been true, despite this epistemological defect. If it had been true, then it is interesting to ask how far stipulative demonstration could have secured determinate denotations for newly introduced terms purporting to refer to electromagnetic properties represented by a potential A^μ. The answer is "Not at all without trivializing theoretical descriptions". The easiest way to see this is to suppose that one were to stipulate that the value of a particular function $A^\mu(\boldsymbol{x}, t)$ at point $(\boldsymbol{x}, t)$ is to count as denoting the actual potential properties associated with that point. No matter for how many points one made such stipulations, the denotation would remain indeterminate for every other point. Only a stipulative demonstration for *every* point could secure determinate denotations for terms purporting to refer to electromagnetic properties represented by a potential A^μ. But if the denotation is rendered determinate in this way, then there will be no room for any empirical statements or thoughts about what properties are associated with what points. The claim that the electromagnetic potential properties associated with a point $(\boldsymbol{x}, t)$ are represented by $A^\mu(x, t)$ would either be trivially true or trivially false. Either way, one could neither have made contingent claims about, nor entertained contentful thoughts about, such properties *even if they had existed*. It is therefore fortunate for the scientific realist that the empirical success of a classical theory of electromagnetism acting on charged particles provides no reason to believe that there are any such properties.

With the advent of quantum theory, it became clear that electromagnetism manifested its effects not simply by altering the dynamics of charged particles, but rather by affecting the relative phases of different components of their associated wave-functions. Aharonov and Bohm (1959) drew attention to the quantum

mechanical prediction that an interference pattern due to a beam of charged particles could be produced or altered by the presence of a constant magnetic field only in a region from which the particles were excluded. The classical force law (7.3) cannot explain this phenomenon: it restricts the dynamic action of electric and magnetic fields to regions where these are nonzero. Classical electromagnetism can explain the effect when combined with quantum mechanics rather than classical mechanics, but the standard explanation appeals not to the field strength $F^{\mu\nu}$, but directly to the potential A^{μ}.[8]

The measurable magnitudes, including now the fringe displacements in interference patterns, are invariant under the gauge transformation (7.2). This is because substitution of A'^{μ} for A^{μ} in the quantum dynamical equation (the Schrödinger equation) with solution ψ results in an equation whose solution is a wave function for the particles with charge q of the form

$$\psi' = \exp\left[-(iq/\hbar)\Lambda(\boldsymbol{x}, t)\right]\psi \tag{7.4}$$

But no measurable magnitude discriminates between the wave functions ψ and ψ'. So a model of the theory incorporating the pair $< \psi, A^{\mu} >$ and an otherwise identical model incorporating instead the pair $< \psi', A'^{\mu} >$ are related by a gauge symmetry that preserves all measurable magnitudes. The lesson is just as in the purely classical case. Formulated in terms of $< \psi, A^{\mu} >$, the theory is multiply realized. This is a symptom of surplus theoretical structure. Understood realistically, the theory is epistemologically defective, because it postulates a theoretical structure that is not measurable even if the theory is true. Models related by a gauge transformation should therefore be taken to represent the same physical situation.

One cannot reformulate this theory of classical electromagnetism in the quantum mechanical context simply by dropping the surplus structure. But that does not affect the central epistemological moral. Even when a realist cannot immediately see how to eliminate surplus structure from the models of a theory which is symmetric under transformations that preserve all measurable structures, (s)he should not take the success of the theory to warrant belief that models related by such a transformation represent distinct situations. The appropriate belief is rather that each equivalence class of models under the relevant symmetry transformations represents the structure of a different situation.

As the present case illustrates, it is not always easy to give an independent description of these situations. The first step is to arrive at a characterization of the structure of models of the reformulated theory that makes it clear what properties and relations are attributed to elements of its domain when it is claimed that a model faithfully represents them. For the realist, the success of the original, epistemologically defective, theory warrants belief in just that structure, however it is described. I have argued elsewhere (Healey (2001)) that the appropriate description of electromagnetic properties in this context is in terms of what I call *holonomy properties*—intrinsic properties of/at closed loops in space(-time) that fail to supervene on intrinsic properties of/at their constituent points. Even if this is correct, it offers little guidance to the realist

on what properties it is reasonable to associate with the charged particles on which these act. But that is just one perspective on the time-honored problem of interpreting quantum mechanics, to whose solution Jeffrey Bub has contributed so much in so many ways.

NOTES

[1] Here there are echoes of the broader issues highlighted by Putnam's Paradox (named and discussed by Lewis in his (1984)), though the "saving constraint" here is provided not by nature but by our demonstrative acts.

[2] Here I am indebted to Tim Maudlin, who suggested a theory like this in correspondence, though I have modified his example for my own purposes.

[3] Indeed, Lewis (1986, 77–78) himself entertained the possibility of such a theory.

[4] Demopoulos and Friedman (1985) take this as an important moral of Newman's (1928) objection to Russell's (1927) structuralism.

[5] *Scholium* to the *Principia*, Newton (1686, [1934, 12]).

[6] See, for example, Friedman (1983) paper III. Sklar (1974) calls the framework neo-Newtonian space-time: Geroch (1978) refers to Galilean space-time.

[7] I have not considered one possible objection to this line of argument suggested by Lewis's (1984) proposed solution to Putnam's Paradox. The idea would be to restrict eligible realizations of a theory to those that "cut Nature at its joints", so that it is Nature itself rather than our causal or intentional links to it, that fixes the reference of our terms—in this case, of the term *is at rest*. Newton might have been sympathetic to such a suggestion, given his own metaphysical views. Or he might have preferred to appeal to God to effect the referential connection, along with his other tasks in constituting space as his sensorium, and periodically restoring the equilibrium of the solar system. I find appeals to metaphysics here no more credible than appeals to theology.

[8] I have discussed this more fully elsewhere (1997, 1999).

REFERENCES

Aharonov, Y. and Bohm, D. (1959), "Significance of electromagnetic potentials in the quantum theory", *Physicl Review 115*: 485–91.

Cruse, P. (2005), "Ramsey sentences, structural realism and trivial realization", *Studies in History and Philosophy of Science 36*: 557–576.

Demopoulos, W. and Friedman, M. (1985), "Critical notice: Bertrand Russell's *The Analysis of Matter*", *Philosophy of Science 52*: 621–639.

Field, H. (1973), "Theory Change and the Indeterminacy of Reference", *Journal of Philosophy 70*: 462–481.

Friedman, M. (1983), *Foundations of Space-Time Theories* (Princeton: Princeton University Press).

Geroch, R. (1978), *General Relativity from A to B* (Chicago: University of Chicago Press).

Healey, R. (1997), "Nonlocality and the Aharonov-Bohm Effect", *Philosophy of Science 64*: 8–41.

——— (1999), "Is the Aharonov-Bohm effect local?", in Cao, T.Y. ed., *Conceptual Foundations of Quantum Field Theory* (Cambridge: Cambridge University Press): 298–309.

——— (2001), "On the reality of gauge potentials", *Philosophy of Science 68*: 432–455.

Ismael, J. and van Fraassen, B.C. (2003), "Symmetry as a guide to superfluous theoretical structure", in Brading, K and Castellani, E. eds., *Symmetries in Physics* (Cambridge: Cambridge University Press): 371–392.

Lewis, David (1970), "How to define theoretical terms", *Journal of Philosophy 67*: 427–446.

——— (1984), "Putnam's paradox", *The Australasian Journal of Philosophy 62*, reprinted in Lewis (1999): 56–77.

Lewis, David (1986), *On the Plurality of Worlds* (Oxford: Blackwell).
——— (1994), "Reduction of mind", in Guttenplan, S. ed., *A Companion to Philosophy of Mind* (Oxford: Blackwell), reprinted in Lewis (1999): 291–324.
——— (1997), "Naming the colors", *The Australasian Journal of Philosophy 75*, reprinted in Lewis (1999): 332–358.
——— (1999), *Papers in Metaphysics and Epistemology* (Cambridge: Cambridge University Press).
——— (forthcoming), "Ramseyan humility", to appear in Braddon-Mitchell, D. and Nola, R. eds., *Naturalistic Analysis* (Cambridge, Mass.: MIT Press).
Newman, M.H.A. (1928), "Mr. Russell's 'Causal theory of perception"', *Mind 37*: 137–148.
Newton, I. (1686, [1934]), *Principia* (Berkeley: University of California Press).
Quine, W.V.O. (1951), "Two dogmas of empiricism", *Philosophical Review 60*: 20–43, reprinted in *From a Logical Point of View* (Cambridge, Mass.: Harvard University Press, 1953).
Redhead, M.L.G. (1975), "Symmetry in intertheory relations", *Synthese 35*: 77–112.
Russell, B. (1927), *The Analysis of Matter* (London: Allen and Unwin).
Sklar, L. (1974), *Space, Time and Spacetime* (Berkeley: University of California Press).
Suppes, P. (1957), *Introduction to Logic* (Princeton: D. Van Nostrand Company, Inc.).
van Fraassen, B.C. (1966), "Singular terms, truth-value gaps, and free logic", *Journal of Philosophical Logic 63*: 481–495.

C.J. ISHAM*

8. IS IT TRUE; OR IS IT FALSE; OR SOMEWHERE IN BETWEEN? THE LOGIC OF QUANTUM THEORY

ABSTRACT

The paper contains a relatively non-technical summary of some recent work by the author and Jeremy Butterfield. The goal is to find a way of assigning meaningful truth values to propositions in quantum theory: something that is not possible in the normal, instrumentalist interpretation. The key mathematical tool is presheaf theory where multi-valued, contextual truth values arise naturally. We show how this can be applied to quantum theory, with the 'contexts' chosen to be Boolean subalgebras of the set of all projection operators.

1 WHAT IS QUANTUM THEORY ABOUT?

Consider the following two statements concerning a physical quantity A and a real number a. The critical words are italicised.

> '*If* a *measurement* of A is made, the probability that the *result* will be a is p.'

> 'The quantity A *has* a value, and the probability that this value *is* a is p.'

The first statement is an instrumentalist way of talking about physics: it does not concern itself with what 'is the case' but only with the results of measurements. The essential counterfactuality is captured by the opening 'If': the statement asserts what would happen (or, more precisely, the probability of what would happen) *if* a certain action is taken. It is silent about the situation in which no measurement is made.

The second statement is very different. It reflects a typical realist view of the world in which, at any moment of time, any physical quantity is deemed to *possess* a value, even if we do not know what that value is. Concomitantly, any proposition asserted about the values of physical quantities is either true or false: a nice, simple, black-or-white view of the world.

In classical physics (and, indeed, in the normal, 'commonsense' world) no fundamental distinction between these statements need be made. If someone asks 'Why

* The Blackett Laboratory, Imperial College of Science, Technology & Medicine, South Kensington, London SW7 2BZ, UK, Corresponding author. Email: c.isham@imperial.ac.uk

W. Demopoulos and I. Pitowsky (eds.), Physical Theory and its Interpretation, 161–182.

did the measurement of the physical quantity A give the particular result that it did?', the obvious answer is that *A possessed* that value at the time the measurement was made. A good measurement simply reveals 'what is the case'.

However, the situation in quantum physics is radically different. The standard interpretation of the theory is unashamedly instrumentalist: indeed, many proponents would insist that it is generally meaningless to even talk about the values of physical quantities other than in the counterfactual language of measurement results.

Quantum theory is usually taught in this way and, of course, within its own limitations the interpretation works extremely well. The rapid growth of the solid-state industries is a striking demonstration of this, as are the activities of the particle physicists at CERN, and those of a host of other scientists and engineers who use quantum theory on a daily basis. However, many scientists (and non-scientists too) feel compelled to seek a deeper reality that lies beneath such an instrumentalist veneer; and even in a strict instrumentalist framework there is still the infamous 'measurement problem' that arises when one probes more deeply into the question of what type of interaction should count as a 'measurement'.

The desire to develop a more realist interpretation of quantum theory reaches an apotheosis in the context of quantum cosmology: the application of quantum theory to the universe itself. However finding such an interpretation is not an easy task, not least because of the difficulty in specifying what is really meant by 'realism' and a 'realist' interpretation. This is, of course, a huge philosophical issue, but in the context of the physical sciences one can tentatively say that a realist interpretation is one in which (i) propositions about the physical world are handled using standard Boolean logic; and (ii) at any moment of time, each such proposition is either true or false. The underlying assumption is that, at any time, every physical quantity *possesses* a definite value. Propositions about the system are then statements that each member of some set of physical quantities has a value that lies in a specific range.

In classical physics, the collection of all propositions about a physical system does indeed form a Boolean algebra (see Section 2); and, for each state of the system, any proposition about the system is indeed either true or false. Of course, all this is in accord with our ordinary, commonsense view of the world.

However, in quantum theory the situation is very different. For example, consider a simple system with a two-dimensional vector space of states, and with the state vector $|\psi\rangle$ shown in Figure 8.1. This could represent the spin degrees of freedom of an electron, with the '$\uparrow$, $\downarrow$' symbols corresponding to the z-component of spin, S_z, being $+\frac{1}{2}\hbar$ and $-\frac{1}{2}\hbar$ respectively.

In the conventional interpretation of quantum theory, all that can be said about the value of S_z is that *if* a measurement is made of the z component of spin, then the *probabilities* of getting the results $-\frac{1}{2}\hbar$ ('down') and $+\frac{1}{2}\hbar$ ('up') are $\cos^2\theta$ and $\sin^2\theta$ respectively (for simplicity I have taken a real, rather than complex, vector space). However, unless $\theta = 0^0$ or 90^0 (so that $|\psi\rangle$ is then an eigenvector of $\hat{S}_z$) nothing can be said about the *value* of the spin: i.e. it cannot be asserted meaningfully that the spin has/possesses any specific value. In particular, the proposition 'the electron has spin down' (or spin up) cannot be assigned a meaningful truth value.

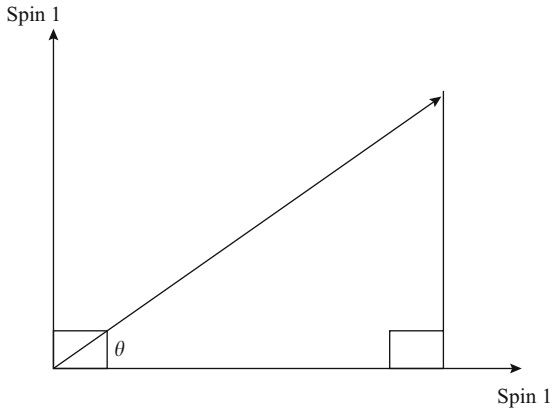

FIGURE 8.1. The quantum state $|\psi\rangle$ is a non-trivial superposition of the two eigenstates ('spin up' and 'spin down') of $\hat{S}_z$. As a consequence, the proposition 'the electron has spin down' has no meaningful truth value.

This inability to sustain a simple realist interpretation of the theory is not just some whimsical psychological preference of the quantum physicist. Rather, it is an inevitable consequence of the famous Kochen–Specker theorem [2]. This asserts the nonexistence[1] of valuations[2] in quantum theory, subject only to the rather plausible requirement that the value of a function of a physical quantity should be the result of applying that function to the value of the quantity. In symbols, if V is a putative value function, and if f is a real-valued function of real numbers then, if A is any physical quantity, the requirement is

$$V(f(A)) = f(V(A)). \tag{1}$$

For example, the value of the quantity 'energy-squared' could reasonably be expected to be the square of the value of the energy.

The Kochen–Specker is a major result in quantum theory, and is the motivational force behind the present paper. When applied to propositions, the theorem asserts the non-existence of any consistent assignment of true-false values to the propositions in quantum theory.

One common response to the Kochen–Specker theorem is to note that although it forbids any absolute assignment of truth values, it does not exclude ones that are *contextual*. Here, 'contextual' means that the truth value given to a proposition depends on which other compatible (meaning 'simultaneously measurable') propositions are given values at the same time. Of course, this does not say how such a contextual valuation might be obtained, or what properties it should possess. The aim of the present paper is to show how one particular such scheme is already contained within the existing formalism of quantum theory, without the need to add hidden variables, or the like. However this scheme has the feature that, as well as being contextual, the

truth values are also *multi-valued*. We shall refer to truth values that are multi-valued and/or contextual as *generalized* truth values.

The idea of multi-valued logic has cropped up from time to time before in the history of quantum theory: for example, Reichenbach introduced the idea of a three-valued logic, so that a proposition could be true, false, or 'in between' [1]. However, a major problem in such proposed logics has always been how to define the logical operations 'and', 'or' and 'not'; in practice, the procedures have tended to be rather hit or miss. A few years ago, Jeremy Butterfield and I introduced a novel form of multi-valued logic in quantum theory that was based on the use of topos theory; or, more precisely, on the use of the special case of presheaf theory. One advantage of this new approach is that the logical operations are defined *unambiguously* by the basic mathematical structure of the relevant presheaf. It is this scheme that is described in the present paper: hopefully, in a relatively non-technical way.

The structure of the paper is as follows. In Section 2 there is a short introduction to the way logic arises in classical physics and in normal quantum theory. This includes a demonstration of how a certain type of multi-valued logic is already present in classical physics. In Section 3 we extend this idea to quantum theory. This involves constructing a special presheaf that can be used to assign truth values that are both contextual and multi-valued. Nevertheless, the underlying logic is sufficiently like that of a Boolean algebra to enable statements about the world to be asserted and manipulated in a logical way.

2 THE LOGIC OF PHYSICS

2.1 The logic of classical physics

A key feature of classical physics is that, at any given time, the system has a definite state, and this state determines—and is uniquely determined by—the values of all the physical quantities associated with the system. The set of possible states of a system is called the 'space of states', or 'state space'. This notion of a state captures well the realist philosophy underlying classical physics.

As an example, consider a point particle moving along a line according to the laws of Newtonian physics. The state of such a system is completely determined by the values of the position, x, and momentum, p, of the particle. Thus the state space is a two-dimensional space with coordinates x and p, as shown in Figure 8.2.

Of course, a point particle has physical properties other than the values of position and momentum; for example, it will have a certain energy, E. However, the energy of the particle is completely determined by its state, i.e. by the values of position and momentum. For example, for a simple harmonic oscillator we have $E(x,p) = (p^2/2m) + kx^2$, where m is the mass of the particle and k is some positive constant.

It is clear that different states can give the same value of the energy. For example, for the harmonic oscillator the set of states (x,p) for which the energy has a value E_1 is represented by the inner ellipse in Figure 8.2. Similarly, the outer ellipse represents the set of states for which the energy has a value E_2, with $E_1 < E_2$. Then the proposition

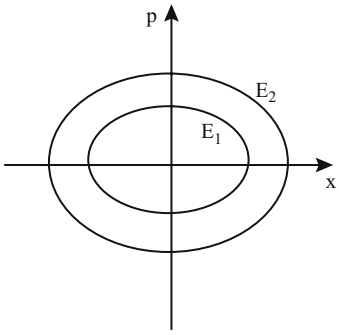

FIGURE 8.2. The classical state space of a particle moving in one dimension. The shaded area represents the set of states for a simple harmonic oscillator for which the energy E satisfies $E_1 < E < E_2$.

'the energy of the system lies between E_1 and E_2' is represented by the yellow subset between the two ellipses.

This idea can be generalized to any classical system. Specifically, if $\mathcal{S}$ is the state space, then every proposition P about the system can be represented by an associated subset, $\mathcal{S}_P$, of $\mathcal{S}$: namely, the set of states for which P is true. Conversely, every subset of $\mathcal{S}$ represents a proposition. More precisely, every subset represents *many* propositions about the values of physical quantities. One sometimes says that two propositions are 'physically equivalent' if they are represented by the same subset of $\mathcal{S}$.

It is easy to see how the logical calculus of propositions arises in this picture. For suppose that P and Q are a pair of propositions, represented by the subsets $\mathcal{S}_P$ and $\mathcal{S}_Q$ respectively, and consider the proposition '*P and Q*'. This is true if, and only if, both P and Q are true, and hence the subset of states representing this logical conjunction consists of those states that lie in both $\mathcal{S}_P$ and $\mathcal{S}_Q$—i.e. the set-theoretic intersection $\mathcal{S}_P \cap \mathcal{S}_Q$. Thus '*P and Q*' is represented by $\mathcal{S}_P \cap \mathcal{S}_Q$. Similarly, the proposition '*P or Q*' is true if either P or Q (or both) are true, and hence this logical disjunction is represented by those states that lie in $\mathcal{S}_P$ plus those states that lie in $\mathcal{S}_Q$—i.e. the set-theoretic union $\mathcal{S}_P \cup \mathcal{S}_Q$. Finally, the logical negation '*not P*' is represented by all those points in $\mathcal{S}$ that do not lie in $\mathcal{S}_P$—i.e. the set-theoretic complement $\mathcal{S}/\mathcal{S}_P$.

In this way, a fundamental relation is established between the logical calculus of propositions about a physical system, and the Boolean algebra of subsets of the state space. Thus the mathematical structure of classical physics is such that, *of necessity*, it reflects a realist philosophy.

2.2 The standard logic of quantum theory

In quantum theory, a proposition is represented [3] by a projection operator[3] on the vector space, $\mathcal{H}$, states. Equivalently, a proposition is represented by the linear subspace, $\mathcal{H}_{\hat{P}}$ (known as the *range* of $\hat{P}$), of $\mathcal{H}$ upon which the projection operator $\hat{P}$ projects. Analogous to the situation in classical physics, many propositions can be

represented by the same projection operator. As we shall see in Section 3, this has important ramifications for what we are trying to do.

If $\hat{P}$ and $\hat{Q}$ are a pair of projection operators, with corresponding subspaces $\mathcal{H}_{\hat{P}}$ and $\mathcal{H}_{\hat{Q}}$ respectively, then the subspace that represents the proposition[4] '*P and Q*' is simply the intersection $\mathcal{H}_{\hat{P}} \cap \mathcal{H}_{\hat{Q}}$: we shall denote the corresponding projection operator by $\hat{P} \wedge \hat{Q}$. Similarly, the subspace that represents the proposition '*not P*' is the orthogonal complement[5] of the subspace $\mathcal{H}_{\hat{P}}$. The corresponding projection operator is $\hat{1} - \hat{P}$, where $\hat{1}$ is the unit operator.

The situation in regard to the logical 'or' operation is more complicated. Given a pair of propositions P, Q, the obvious choice to represent '*P or Q*' might seem to be the union $\mathcal{H}_{\hat{P}} \cup \mathcal{H}_{\hat{Q}}$. However, this is not a linear subspace of the vector space $\mathcal{H}$, and hence cannot represent any proposition. Instead, the proposition '*P or Q*' is represented by the *linear span* of the vectors in $\mathcal{H}_{\hat{P}} \cup \mathcal{H}_{\hat{P}}$—i.e. the collection of all possible sums of vectors in $\mathcal{H}_{\hat{P}} \cup \mathcal{H}_{\hat{P}}$; the corresponding projection operator will be denoted by $\hat{P} \vee \hat{Q}$. This choice has the desirable property of *associativity*: for any three projectors $\hat{P}$, $\hat{Q}$ and $\hat{R}$ we have $\hat{P} \vee (\hat{Q} \vee \hat{R}) = (\hat{P} \vee \hat{Q}) \vee \hat{R}$. This is consonant with the logic of daily life where it is taken for granted that if P, Q, R are any three propositions, then '(*P or Q*) *or R*' = '*P or* (*Q or R*)'. It is easy to see that the 'and' operation is also associative: for any three projectors $\hat{P}, \hat{Q}, \hat{R}$, we have $\hat{P} \wedge (\hat{Q} \wedge \hat{R}) = (\hat{P} \wedge \hat{Q}) \wedge \hat{R}$.

However, this 'quantum logic' of projection operators differs from Boolean logic in one critical feature: it fails to be distributive. Thus, given three projectors $\hat{P}, \hat{Q}, \hat{R}$, we will generally have

$$\hat{P} \wedge (\hat{Q} \vee \hat{R}) \neq (\hat{P} \wedge Q) \vee (\hat{P} \wedge R). \tag{2}$$

To see how bizarre non-distributive thinking would be in daily life suppose that I was staying at a hotel and, at breakfast, the waiter said 'Would you like eggs and sausage or bacon?'. If I parsed this phrase as 'eggs and (sausage or bacon)', I would assume that I was being offered a choice between eggs and sausage, or eggs and bacon. In other words, I would invoke the distributive law

$$E \textit{ and } (S \textit{ or } B) = (E \textit{ and } S) \textit{ or } (E \textit{ and } B). \tag{3}$$

However, it is easy to construct a simple quantum model in which if I respond 'eggs and sausage, please' I get nothing, and similarly for eggs and bacon. In fact, in this particular example, the only sensible reply to 'Would you like eggs and sausage or bacon?' is 'Yes please', in which case my plate would arrive with eggs plus a quantum superposition of sausage and bacon.

As applied to this non-distributive quantum logic, the Kochen–Specker theorem asserts the impossibility of assigning consistent true-false values to projection operators. Thus the corresponding properties cannot be said to be 'possessed' by the system. However, the Kochen–Specker theorem does not preclude the existence of truth

values that are contextual and/or multi-valued, provided an appropriate mathematical structure can be found. It is to this task that we now turn.

2.3 A role for multi-valued logic in classical physics

Consider first a classical system with state space $\mathcal{S}$. Each physical quantity A is represented by a real-valued function (denoted $\overline{A}$) on $\mathcal{S}$ with the interpretation that if s in $\mathcal{S}$ is a state of the system, then the value of the physical quantity A in that state is the real number $\overline{A}(s)$. As explained in Section 2.1, a proposition of the form '$A \in \Delta$' (meaning that the value of A lies in the set Δ of real numbers) is then represented by the subset $\mathcal{S}_{A\in\Delta}$ of $\mathcal{S}$ consisting of all those states s for which $\overline{A}(s)$ belongs to Δ (see Figure 8.3). Of course, this structure is consistent with the philosophical view that each physical quantity *has* a value for any given state of the system. In particular, any proposition asserted about the system is either true or false. Thus the proposition '$A \in \Delta$' is true if s belongs to $\mathcal{S}_{A\in\Delta}$, and it is false if it does not.

All this seems clear-cut—but is it really so? For suppose s is a state that does not belong to $\mathcal{S}_{A\in\Delta}$ but which, nevertheless, is 'almost' in this subset (so that $\overline{A}(s)$ 'almost' belongs to Δ): is there not some sense in which the proposition '$A \in \Delta$' is then 'almost true'? Contrariwise, suppose s is such that $\overline{A}(s)$ belongs to Δ, but only just so (i.e. $\overline{A}(s)$ is 'close' to the edges of Δ): then is '$A \in \Delta$' not 'almost false', or 'only just true'? Such grey-scale judgements are frequently made in daily life, but there seems to be no role for them in the harsh, black-and-white mathematics of classical physics.

The situation becomes even more piquant if, rather than being given a specific (micro)-state s, we know only that s lies in some subset M (a *macro-state*) of $\mathcal{S}$.[6] What truth value, if any, can then be ascribed to the proposition '$A \in \Delta$'? If M is a subset of $\mathcal{S}_{A\in\Delta}$, it does seem correct to say that the proposition is true (perhaps even 'totally true'?), since for each state s in the macro-state M, the real number $\overline{A}(s)$ *does* belong to Δ.[7]

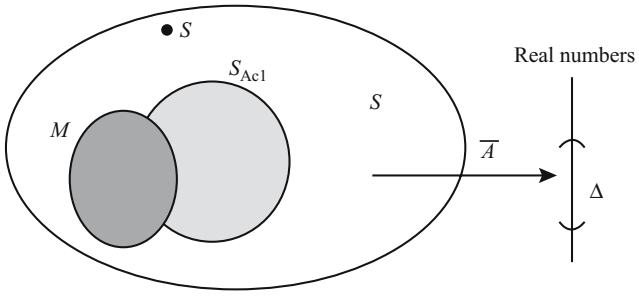

FIGURE 8.3. A diagram to aid discussing a multi-valued logic for classical macrostates. Here $\mathcal{S}$ is the state-space for the system; $M \subset \mathcal{S}$ is a macrostate; $\overline{A}$ is the real-valued function on $\mathcal{S}$ that represents a physical quantity A; and $\mathcal{S}_{A\in\Delta}$ is the subset of states s such that $\overline{A}(s)$ belongs to $\Delta \subset \mathbb{R}$.

However, suppose M is such that there are some states s in M for which $\overline{A}(s)$ belongs to Δ, and some states for which it does not (see Figure 8.3). In this situation, what truth value can be given to the proposition '$A \in \Delta$'? Is there some sense in which it is 'partially true'?

As things stand, if '$A \in \Delta$' is interpreted as asserting that, for *all* s in M, the real number $\overline{A}(s)$ belongs to Δ, then the proposition is clearly false. But suppose M is 'almost' a subset of $\mathcal{S}_{A\in\Delta}$: are we not then tempted to say that the proposition '$A \in \Delta$' is 'almost true'? At the very least, in such circumstances it seems misleading to assert that the proposition is unequivocally false. And even if none of the states in M belong to $\mathcal{S}_{A\in\Delta}$ we can still imagine situations in which M is 'close' to this subset, so that it might appropriate to say that '$A \in \Delta$' is 'almost true' (or, perhaps, 'almost not false').

The difficulty in succumbing to such temptations is that the word 'almost' has no well-defined meaning in the standard mathematics that is used in physics. But what is strongly suggested by the discussion above is the need to introduce multiple truth values that can interpolate between 'true' and 'false'.

2.4 The use of coarse-graining

One way of introducing multi-valued logic into classical physics is to use a certain coarse-graining operation. The basic idea is rather simple. Namely, if we are in the type of situation envisaged above, where we feel reluctant to assign a simple true-false value to a proposition '$A \in \Delta$', then perhaps we can find a real-valued function f of the real numbers such that the proposition[8] '$f(A) \in f(\Delta)$' definitely *is* true. This possibility arises because the proposition '$f(A) \in f(\Delta)$' is *weaker* than the proposition '$A \in \Delta$': thus '$A \in \Delta$' *implies* '$f(A) \in f(\Delta)$' but the converse is generally not true. For example from the knowledge that a physical quantity A has the value 2, the quantity A^2 can be deduced to have the value 4. On the other hand, from the knowledge that $A^2 = 4$ all that can be deduced about the value of A is that it is equal to $+2$ or -2. This weakening of propositions can occur whenever the function f is not one-to-one.

The question now is if such weakening operations can be used to give a truth value to a proposition '$A \in \Delta$' in a macro-state M when M is not simply a subset of $\mathcal{S}_{A\in\Delta}$—i.e. $\overline{A}(M)$ is not a subset of Δ. As discussed in Section 2.3, in this situation we may be reluctant to say that '$A \in \Delta$' is just false.

Our, at first rather implausible looking, suggestion is that the generalized truth value of the proposition '$A \in \Delta$' in the macro-state M is to be related to the set of all functions f such that $f(\overline{A}(M))$ *is* a subset of $f(\Delta)$. This condition can be rewritten as $f \circ \overline{A}(M) \subset f(\Delta)$, where $f \circ \overline{A}$ is the function from $\mathcal{S}$ to $\mathbb{R}$ defined by $f \circ \overline{A}(s) := f(A(s))$ for all s in $\mathcal{S}$. We shall denote by $f(A)$ the physical quantity corresponding to the function $f \circ \overline{A}$, and say that $f(A)$ is a *coarse-graining* of A. Then the precise form of our suggestion is that the generalized truth value, $V^M(A \in \Delta)$ of the proposition '$A \in \Delta$' is to be *defined* as the set of all coarse-grainings, $f(A)$, of A for which the weaker proposition '$f(A) \in f(\Delta)$' *is* true, in the usual sense of 'true'!

In symbols, we define the generalized valuation

$$V^M(A \in \Delta) := \{f \mid f(\overline{A}(M)) \subset f(\Delta)\}. \tag{4}$$

Of course, it is not obvious that truth values defined in this way have a logical structure; but they do! The proof involves presheaf theory: a subject which we will introduce shortly in the context of finding generalized truth values in quantum theory. Understandably, this involves coarse-graining *operators*, since it is these that represent physical quantities in quantum theory. The use of presheaf ideas in classical physics is discussed in [4].

3 THE PRESHEAF LOGIC OF QUANTUM THEORY

3.1 Coarse-graining in a quantum context

The standard, instrumentalist interpretation of quantum theory gives the probability that a proposition '$A \in \Delta$' will be found to be true if measurements[9] are made of the physical quantity A. Specifically, if $|\psi\rangle$ is a normalized state, the probability that the results will lie in the subset Δ of real numbers is

$$\text{Prob}(A \in \Delta;\, |\psi\rangle) = \langle\psi|\, \hat{E}[\, A \in \Delta\,]\, |\psi\rangle, \tag{5}$$

where $\hat{E}[\, A \in \Delta\,]$ denotes the projection operator onto the subspace of eigenvectors of $\hat{A}$ whose eigenvalues lie in Δ. The operator $\hat{E}[\, A \in \Delta\,]$ is known as a *spectral projector* of the operator $\hat{A}$ that represents the physical quantity A.

A more realist interpretation might aspire to give a truth value to the proposition '$A \in \Delta$' without invoking external measurements. If the quantum state $|\psi\rangle$ is such that $\text{Prob}(A \in \Delta;\, |\psi\rangle) = 1$, it is arguably meaningful to assert that '$A \in \Delta$' is true. Contrariwise, if $\text{Prob}(A \in \Delta;\, |\psi\rangle) = 0$ it might seem natural to say that '$A \in \Delta$' is false, although—motivated by the discussion in Section 2.3 of classical macrostates—one might want to think about situations in which $|\psi\rangle$ is 'close' to a state for which the probability of '$A \in \Delta$' is greater than zero[10]. In any event, in the cases where $0 \leq \text{Prob}(A \in \Delta;\, |\psi\rangle) < 1$ it is certainly not the case that '$A \in \Delta$' is simply true.

One approach would be to define the truth value of '$A \in \Delta$' to *be* the probability $\text{Prob}(A \in \Delta;\, |\psi\rangle)$. This involves the use of *fuzzy logic* in which the truth values of propositions are real numbers in the interval $[0, 1]$. However, we shall adopt a different tack by invoking an operator analogue of the coarse-graining operations used in Section 2.4 in the context of classical physics.

One of the basic structural assumptions in quantum theory is that for any function f, the operator that represents the coarse-grained physical quantity $f(A)$ is $f(\hat{A})$: in this sense, $f(\hat{A})$ is a 'coarse-graining' of the operator $\hat{A}$. Additionally, it is easy to show that the spectral projectors $\hat{E}[\, A \in \Delta\,]$ and $\hat{E}[f(A) \in f(\Delta)\,]$ satisfy $\hat{E}[\, A \in \Delta\,] \preceq \hat{E}[f(A) \in f(\Delta)\,]$, where $\hat{P}_1 \preceq \hat{P}_2$ denotes that $\hat{P}_1$ projects onto a subspace of the range of $\hat{P}_2$ (i.e. $\mathcal{H}_{P_1}$ is a subspace of $\mathcal{H}_{P_2}$). In this sense, the projection operator $\hat{E}[f(A) \in f(\Delta)\,]$ is a coarse-graining of $\hat{E}[\, A \in \Delta\,]$.

Guided by the discussion of classical physics in Section 2.4, one suggestion might be that, for any given quantum state $|\psi\rangle$, the generalized truth value of the proposition '$A \in \Delta$' is the collection of coarse-grainings, $f(\hat{A})$, of $\hat{A}$ such that the weaker proposition '$f(A) \in f(\Delta)$' *is* true—i.e. it is true with probability one, so that $\langle\psi|\hat{E}[f(A) \in f(\Delta)]|\psi\rangle = 1$. In other words, we could try the definition (cf. equation (4))

$$V^{\psi}(A \in \Delta) := \{f \mid \langle\psi|\hat{E}[f(A) \in f(\Delta)]|\psi\rangle = 1\}. \tag{6}$$

This possibility arises since $\hat{E}[A \in \Delta] \preceq \hat{E}[f(A) \in f(\Delta)]$ implies that, for any f,

$$\langle\psi|\hat{E}[A \in \Delta]|\psi\rangle \leq \langle\psi|\hat{E}[f(A) \in f(\Delta)]|\psi\rangle \tag{7}$$

for all quantum states $|\psi\rangle$. Hence, even if $\langle\psi|\hat{E}[A \in \Delta]|\psi\rangle < 1$, there can be functions f such that $\langle\psi|\hat{E}[f(A) \in f(\Delta)]|\psi\rangle = 1$.

The use of equation (6) is perfectly viable, and is discussed in detail in [4]. This includes the construction of the appropriate presheaf needed to show that collections of functions of the type in the right-hand side of equation (6) do have a logical structure.

However, we shall proceed here in a somewhat different way in order to bring out the connection with standard quantum logic. In particular, as explained in Section 2.2, the logical operations 'and', 'or' and 'not', are defined on *projection operators*, not on the underlying propositions. Similarly, the Kochen–Specker theorem deals with the existence of true-false valuations on projectors, not propositions *per se*. This suggests that we should work *ab initio* with projection operators, and hence consider the generalized valuation

$$V^{\psi}(\hat{E}[A \in \Delta]) := \{f \mid \langle\psi|\hat{E}[f(A) \in f(\Delta)]|\psi\rangle = 1\}. \tag{8}$$

At this point an important subtlety arises. Namely, it is possible for a pair of propositions '$A \in \Delta$' and '$B \in \Delta'$' to be represented by the *same* projection operator:

$$\hat{E}[A \in \Delta] = \hat{E}[B \in \Delta'] \tag{9}$$

even if the corresponding operators $\hat{A}$ and $\hat{B}$ do not commute[11]. But then, letting $\hat{P}$ denote $\hat{E}[A \in \Delta] = \hat{E}[B \in \Delta']$, equation (8) gives the two generalized valuations

$$V^{\psi}(\hat{P}) := \{f \mid \langle\psi|\hat{E}[f(A) \in f(\Delta)]|\psi\rangle = 1\}, \tag{10}$$

$$V^{\psi}(\hat{P}) := \{g \mid \langle\psi|\hat{E}[g(B) \in g(\Delta')]|\psi\rangle = 1\} \tag{11}$$

and there is no reason why equation (10) and equation (11) should be equal.

Propositions of this type arise when an operator $\hat{O}$ has vanishing commutators with a pair of operators $\hat{C}, \hat{D}$ with $[\hat{C}, \hat{D}] \neq 0$. For example, let $\hat{O}$ be the Hamiltonian $\hat{H}$ of the hydrogen atom, and let $\hat{C}$ and $\hat{D}$ be $\hat{L}_x$ and $\hat{L}_y$—the x and y components of angular

momentum respectively. Then $[\hat{H}, \hat{L}_x] = 0 = [\hat{H}, \hat{L}_y]$, and $[\hat{L}_x, \hat{L}_y] = i\hbar\hat{L}_z \neq 0$. Now, the spectral theorem for commuting operators asserts the existence of hermitian operators $\hat{A}$ and $\hat{B}$ such that $\hat{H}$ and $\hat{L}_x$ are functions of $\hat{A}$, and $\hat{H}$ and $\hat{L}_y$ are functions of $\hat{B}$. Thus, for some set of functions f, g, h, k we have

$$\hat{H} = f(\hat{A}),\ \hat{L}_x = g(\hat{A}), \tag{12}$$

$$\hat{H} = h(\hat{B}),\ \hat{L}_y = k(\hat{B}). \tag{13}$$

Then, for any[12] subset J of the real numbers, we have[13] $\hat{E}[H \in J] = \hat{E}[f(A) \in J] = \hat{E}[A \in f^{-1}(J)]$, and similarly $\hat{E}[H \in J] = \hat{E}[h(B) \in J] = \hat{E}[B \in h^{-1}(J)]$. Thus $\hat{E}[A \in f^{-1}(J)] = \hat{E}[B \in h^{-1}(J)]$, and of course $[\hat{A}, \hat{B}] \neq 0$ since $[\hat{L}_x, \hat{L}_y] \neq 0$. Hence this provides an example of the situation envisaged above in regard to equation (9), with Δ and Δ' chosen to be $f^{-1}(J)$ and $h^{-1}(J)$ respectively.

What this discussion implies is that the truth value assigned to a projection operator $\hat{P}$ should be *contextual*, i.e. it depends on the physical quantity with which one thinks of $\hat{P}$ as being associated. In the example above of the hydrogen atom, with $\hat{P}$ chosen as $\hat{E}[H \in J]$, the choice is between thinking of this projector as being associated with $\hat{A}$, or with $\hat{B}$. Equivalently, the truth value assigned to the proposition '$H \in J$' depends on whether H is thought of in the context of simultaneously ascribing a truth value to propositions about L_x, or to propositions about L_y.

That such ideas should enter at this point is not surprising since, as remarked earlier, discussions of the physical implications of the Kochen–Specker theorem frequently introduce the notion of contextuality. The important question now is to decide on the most appropriate mathematical framework in which to explore the implications of equation (10) and equation (11). There are, in fact, several different (but mathematically equivalent) approaches to this issue, depending on what one decides to call a 'context'.

As indicated above, one choice is to define a context for a projection operator $\hat{P}$ as one of the operators for which it is a spectral projector: this means using the definition in equation (6) for all physical quantities A and subsets Δ for which $\hat{E}[A \in \Delta] = \hat{P}$. The mathematical development of this idea involves re-expressing equation (8) in the language of presheaf theory and is discussed in [5].

Another possibility is to define a context as an algebra of simultaneously commuting operators to which $\hat{P}$ belongs. In the example of the hydrogen atom, if $\hat{P} = \hat{E}[H \in J]$, then two such algebras are those generated by $\hat{H}$ and $\hat{L}_x$, and by $\hat{H}$ and $\hat{L}_y$, respectively. This approach is discussed in [6]. However, since the focus of the present paper is logic, we shall use a third possibility, which is to define a context for a projector $\hat{P}$ as a *Boolean algebra* to which $\hat{P}$ belongs. The details are as follows.

3.2 Windows on reality

For each hermitian operator $\hat{A}$, let W_A denote the collection of all projection operators of the form $\hat{E}[A \in \Delta]$, as Δ ranges over the subsets of the real numbers.

This forms a Boolean subalgebra of the non-distributive algebra $\mathcal{L}$ of all projection operators. In addition, for any function f, we have $\hat{E}[f(A) \in f(\Delta)] = \hat{E}[A \in f^{-1}(f(\Delta))]$, and hence $\hat{E}[f(A) \in f(\Delta)]$ belongs to W_A. In this sense, equation (10) and equation (11) can be said to assign truth values to the projection operator $\hat{P} = \hat{E}[A \in \Delta] = \hat{E}[B \in \Delta']$ in the *context* of W_A and W_B respectively. In the example of the hydrogen atom, W_A will include the spectral projectors of $\hat{H}$ and $\hat{L}_x$, and W_B will include the spectral projectors of $\hat{H}$ and $\hat{L}_y$.[14]

These remarks suggest that the set, $\mathcal{W}$, of all Boolean sub-algebras of $\mathcal{L}$ is a possible space of contexts in which to assert generalized truth values of projection operators. I shall refer to each such Boolean sub-algebra as a *window* since it gives a partial, Boolean view of the quantum world: a Boolean sub-algebra provides a 'window on reality'.

The next step is to explore the mathematical structure of $\mathcal{W}$. A key property is that it is a partially-ordered set if an ordering (denoted $<$) between windows W_1, W_2 is defined by[15]

$$W_1 < W_2 \text{ if } W_2 \subset W_1, \tag{14}$$

where the right-hand side is to be read as saying that W_2 is a Boolean sub-algebra of W_1 (not just a subset). That this is a partial-ordering[16] is easy to check. From a logical perspective, if $W_2 \subset W_1$ then every element in W_2 can be written as the logical 'or' of disjoint elements in W_1, and hence if $W_1 < W_2$, one can say that W_2 is a *coarse-graining* of W_1. This is represented in Figure 8.4 where the subsets of the

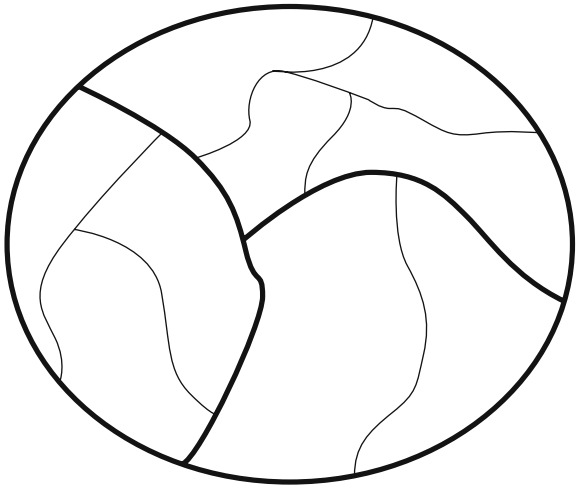

FIGURE 8.4. Representation of a situation in which two windows W_1, W_2 satisfy $W_1 < W_2$. The thin lines symbolically enclose elements of the Boolean algebra W_1; the thick lines enclose elements of W_2 that are coarse-grainings of the elements of W_1.

plane bounded by the red and green lines represent the disjoint projectors in W_1 and W_2 respectively.

This ordering on windows is consistent with the $\preceq$-ordering on projection operators in the sense that, for any coarse-graining $f(\hat{A})$ of $\hat{A}$, we have (i) $\hat{E}[A \in \Delta] \preceq \hat{E}[f(A) \in f(\Delta)]$; and (ii) $W_{f(A)} \subset W_A$, i.e. $W_A < W_{f(A)}$.

3.3 The presheaf of local truth values

Because of the Kochen–Specker theorem, binary truth values cannot be assigned consistently to $\mathcal{L}$. However they *can* be assigned to any of its Boolean sub-algebras, W. Such a valuation is a homomorphism from W to the simplest Boolean algebra $\{0, 1\}$, with 0 and 1 being interpreted as 'false' and 'true' respectively.

The next step is to associate with each window W the set, D_W, of all valuations on W. A crucial observation is that if $W_1 < W_2$, there is a map $k_{W_1W_2}$ from D_{W_1} to D_{W_2}. Specifically, let χ be a valuation on W_1: then, since W_2 is a subalgebra of W_1, we can define χ on W_2 by using the values it assign to elements of W_2 considered as members of W_1. These maps $k_{W_1W_2}$ have the property that if $W_1 < W_2 < W_3$ then

$$k_{W_1W_3} = k_{W_2W_3} \circ k_{W_1W_2}. \tag{15}$$

This means we have an example of a *presheaf* on the partially-ordered set ('poset') $\mathcal{W}$ of windows. We shall call it the 'presheaf of local truth values'.

To introduce the definition of a presheaf it may be helpful to contrast it with the simpler concept of a *fibre bundle*—something that is much used in modern theoretical physics. A fibre bundle with base space[17] B is an association to each point b in B of a space F_b (the 'fibre over b') with the property that these fibres are all copies of a single space F, known as 'the fibre' of the bundle. The 'bundle space' is then defined to be the union of all the fibres F_b, b in B.

The simplest example of a fibre bundle is a product bundle, defined to be the set of all pairs (b, v) where b is in B, and v in F. Bundles of this type are called 'trivial'. An example is given in Figure 8.5 where the base space is a circle, S^1, and the fibre has just two points. Figure 8.6 is an example of a non-trivial bundle with the same fibre and base space. This can be thought of as a Möbius strip with everything but the edges of the strip removed. We see that the fibres 'twist' around as we move round the base space.

An important idea in fibre-bundle theory is that of a *cross-section*. This is defined to be a continuous function from the base space B to the bundle space, with the property that each point b in B is mapped to some point in the fibre F_b over b[18]. For the trivial bundle in Figure 8.5 there are just two cross-sections, corresponding to mapping the base space circle into the lower, and upper, circles in the bundle space respectively.

For non-product bundles the situation is different, and there may be no cross-sections at all. For example, this is true of the bundle in Figure 8.6: any attempt to construct a continuous cross-section inevitably leads to a discontinuity as one works around the base space and comes back to the starting point.

After this preamble, we can return to the idea of a presheaf. A *presheaf*[19] X over a poset $\mathcal{P}$ is defined to be (i) an association to each p in $\mathcal{P}$ of a space X_p (known as

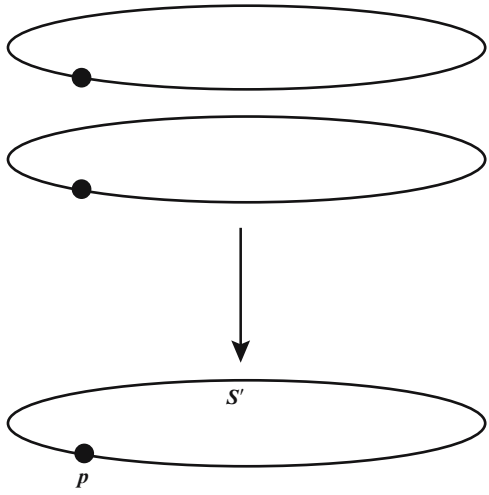

FIGURE 8.5. A trivial fibre bundle whose base space is a circle, and whose fibre over each point is a set with two elements.

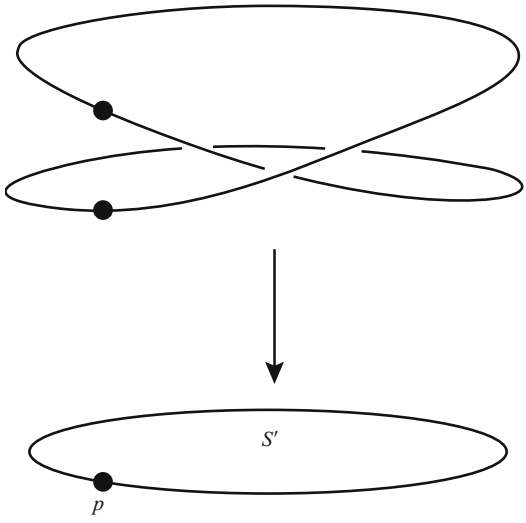

FIGURE 8.6. A non-trivial fibre bundle with the same base space as in Figure 8.5. Each fibre is again a set with two elements, but the bundle has a non-trivial 'twist' and is not the same as the bundle space in.

the *stalk* over p); and (ii) an association to each pair p, q such that $p < q$, of a map X_{pq} from X_p to X_q that satisfies the 'coherence' condition that if $p < q < r$ then (c.f. equation (15))[20]

$$X_{pr} = X_{qr} \circ X_{pq}. \tag{16}$$

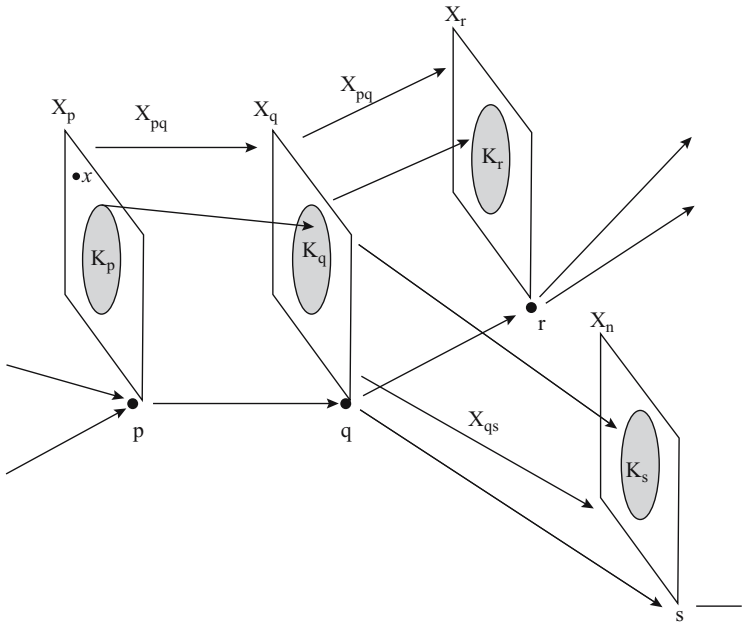

FIGURE 8.7. A presheaf X on a partially ordered set $\mathcal{P}$. This is an association of (i) to each point p in $\mathcal{P}$, a set X_p; and (ii) to each pair of points p, q in $\mathcal{P}$ such that $p < q$, a map X_{pq} from X_p to X_q. These maps satisfy the coherence condition that if $p < q < r$ then $X_{pr} = X_{qr} \circ X_{pq}$. The diagram also illustrates a sub-presheaf K of X.

This is illustrated in Figure 8.7, where the letters p, q, r, s denote elements in $\mathcal{P}$, and the notation $p \longrightarrow q$ means that $p < q$. Thus a presheaf resembles a fibre bundle except that (i) the stalks at different points in $\mathcal{P}$ need not be copies of a single space (unlike the fibres in a fibre bundle); and (ii) the maps X_{pq} exist when $p < q$.

There is much more to this structure than meets the eye. In particular, presheaf theory can be viewed as a *generalization* of set theory itself! Specifically, a single set gets replaced by a *parameterised* family of sets X_p, p in $\mathcal{P}$, that are related by the maps X_{pq} from X_p to X_q; this is why a presheaf is sometimes known as a 'varying set'. This generalization of set theory is very important and is an important part of topos theory [7]. As we shall see, a presheaf embodies a generalization of the Boolean logic of normal set theory.

Various standard ideas in set theory can be extended to this new context. For example, the analogue of an element x of a set X is a *global element* x of a presheaf X. This is defined to be an association to each p in $\mathcal{P}$ of a point x_p in X_p such that if $p < q$ then $x_q = X_{pq}(x_p)$. A related idea is a 'partial' element, where the points x_p are defined over only some sub-poset of $\mathcal{P}$. Thus a global element of a presheaf resembles a cross-section of a fibre bundle, and a partial element resembles a bundle section defined on some subset of the base space.

It is not difficult to show that the Kochen–Specker theorem is equivalent to the statement that the presheaf of local truth values has no global elements (although there are partial elements). This is reminiscent of the result that a certain type of fibre bundle has no cross-sections if the bundle is non-trivial; an example is the bundle in Figure 8.6. Thus we can think of the Kochen–Specker theorem as saying that the presheaf of local truth values is 'twisted' as we move around the space $\mathcal{W}$ of all windows, rather as in Figure 8.6 the fibres twist around the circle S^1!

3.4 The presheaf origin of contextual truth values

Another crucial set-theoretic concept with a presheaf analogue is that of a subset. A *sub-presheaf* (the analogue of a subset) is defined to be an assignment to each p in $\mathcal{P}$ of a subset K_p of X_p with the property that if $p < q$ then $X_{pq}(K_p) \subset K_q$ (cf. Figure 8.7). As we shall see, this concept leads to a powerful mathematical way of encoding the idea of 'contextual truth'.

The connection with logic arises in the following way. First consider normal set theory. Then any subset K of a set X is uniquely specified by its characteristic function χ^K on X, defined by

$$\chi^K(x) = \begin{cases} 1 & \text{if } x \text{ is in } K, \\ 0 & \text{otherwise} \end{cases} \tag{17}$$

for all x in X. To each subset K, and to each point x in X, there is an associated proposition '$x \in K$', and we can think of χ^K as a valuation of these propositions. Then equation (17) asserts that the proposition '$x \in K$' is true if and only if x belongs to K: not a terribly surprising result!

However, the presheaf analogue is more subtle, with truth values now being both contextual and multi-valued. This is reflected in the presheaf analogue of a characteristic function. First we need to define the appropriate analogue of the simple set $\{0, 1\}$ of truth values in standard set theory. This involves introducing the concept of a 'sieve at a point p' in $\mathcal{P}$. This is defined to be a subset S of $\mathcal{P}$ such that (i) $p < q$ for all q in S; and (ii) if q belongs to S and $q < q'$, then q' belongs to S. In the quantum case, a sieve on a window W is a collection of coarse-grainings of W such that if some W' belongs to the collection, then so does any coarse-graining of W'.

One of the fundamental results of presheaf theory is that the collection of all sieves at a point p has the structure of a *logic*. Specifically: the operations of 'and' and 'or' on a pair of sieves S_1, S_2 on p are defined to be the intersection $S_1 \cap S_2$ and the union $S_1 \cup S_2$, respectively. These operations are associative and distributive. The operation of negation is more complicated since, if S is a sieve on p, the obvious guess for $\neg S$—the complement of S—is not itself a sieve. Instead, $\neg S$ is the subset of $\mathcal{P}$ defined by

$$\neg S := \{q \mid p < q \text{ and for all } r \text{ with } q < r, r \text{ does not belong to } S\} \tag{18}$$

which *is* a sieve. With this definition, the collection of all sieves on a point p in $\mathcal{P}$ has the structure of a logic that is almost Boolean, where 'almost' is understood in the following sense.

A Boolean algebra satisfies the famous principle of the excluded middle: namely, a proposition P is always either true or false—in mathematical terms, $P \vee \neg P = 1$, where 1 denotes the proposition that is identically true. For sieves, however, $P \vee \neg P$ is not identically true. Logics of this type are known as *Heyting algebras* and have been much studied in the topos literature. They can be used to form *deductive systems*, and are hence a bona fide alternative to the familiar logic of daily life.

The relation between sieves and a sub-presheaf is as follows. Let a collection of sets $K_p \subset X_p$, p in $\mathcal{P}$, be a sub-presheaf K of the presheaf X. Then there is an associated characteristic 'arrow', which, for each p in $\mathcal{P}$, is a map $\chi_p^{\mathbf{K}}$ from X_p to the set of sieves on p, defined by

$$\chi_p^{\mathbf{K}}(x) := \{q \mid p < q \text{ and } X_{pq}(x) \text{ is in } K_q\} \tag{19}$$

for all x in X_p (Figure 8.7 may help at this point). Thus, at p in $\mathcal{P}$, the characteristic arrow assigns to any element x in X_p the set of all points q in $\mathcal{P}$ for which the transformed point $X_{pq}(x)$ *does* belong to the subset K_q of X_q. The definition of a sub-presheaf then implies that the right-hand side of equation (19) is a *sieve* on p.

3.5 The coarse-graining presheaf

When X is a classical state space $\mathcal{S}$, the definition of a characteristic function in equation (17) provides another way of understanding why the logic of classical physics is Boolean, and with each proposition being either true or false. For, as discussed in Section 2.1, a proposition of the form '$A \in \Delta$' is represented by the subset $\mathcal{S}_{A\in\Delta}$ of $\mathcal{S}$. Then the characteristic function of $\mathcal{S}_{A\in\Delta}$ assigns to a state s the values 1 ('true') if $\overline{A}(s)$ is in Δ, and 0 ('false') if $\overline{A}(s)$ is not in Δ.

On the other hand, as explained in Section 3.1, our suggestion in quantum theory is that the truth value associated with a projection operator $\hat{P}$ should be *contextual*, in the sense that it depends on the window to which one thinks of $\hat{P}$ as belonging. The discussion surrounding equation (19) then suggests that we might try to define the generalized truth value of a projector $\hat{P}$ in the context of a window W (where $\hat{P}$ belongs to W) to be a sieve of windows on W associated with a sub-presheaf of some presheaf on $\mathcal{W}$.

We might anticipate that this construction should be connected in some way with the type of coarse-graining operation discussed in Section 3.1 in which a projector $\hat{E}[A \in \Delta]$ is replaced with $\hat{E}[f(A) \in f(\Delta)]$. However, at the moment, the only presheaf we have is the presheaf of local truth values, which, since it does not involve coarse-graining, cannot serve for our present purposes.

To proceed further, we return to the ordering operation on $\mathcal{W}$, in which $W_1 < W_2$ means that $W_2 \subset W_1$, and ask how coarse-graining might enter here. More precisely, if $\hat{P}$ is a projection operator in W_1, can it be associated with a projection operator, to be denoted $G_{W_1W_2}(\hat{P})$, that belongs to W_2 and is such that $\hat{P} \preceq G_{W_1W_2}(\hat{P})$?

In fact, typically there are many such projectors in W_2 (Figure 8.4 may help here), and it is natural to choose the one that is the 'best approximation' to $\hat{P}$, meaning the 'smallest' (the infinum with respect to the $\preceq$-ordering) projector $\hat{Q}$ in W_2 such that $\hat{P} \preceq \hat{Q}$. Hence we define, for all $\hat{P}$ in W_1,

$$G_{W_1W_2}(\hat{P}) := \inf\{\hat{Q} \text{ in } W_2 \mid \hat{P} \preceq \hat{Q}\}. \tag{20}$$

Thus $G_{W_1W_2}$ is a map from the Boolean algebra W_1 to the Boolean algebra W_2, and it can be shown that, if $W_1 < W_2 < W_3$, then (cf. equation (15) and equation (16))

$$G_{W_2W_3} \circ G_{W_1W_2} = G_{W_1W_3}. \tag{21}$$

It follows that we have constructed a presheaf, denoted G, over $\mathcal{W}$ in which (i) the stalk associated with each window W is a copy of this Boolean algebra, i.e. $G_W := W$; and (ii) the map $G_{W_1W_2}$ from $G_{W_1} = W_1$ to $G_{W_2} = W_2$ is defined by equation (20).

We shall call G the *coarse-graining presheaf*, and use it to assign contextual, multi-valued truth values to quantum propositions. The intention is to exploit the notion of a characteristic arrow defined in equation (19), and, in particular, the fundamental result that the right-hand side of this equation is a sieve. Thus the final step is to find the appropriate sub-presheaf of G and, inspired by equation (8), we proceed as follows.

For each quantum state $|\psi\rangle$, and each context W, we define the set of 'totally true' projectors in W to be the subset, T^{ψ}_{W}, of projectors $\hat{Q}$ in the Boolean algebra W such that $\langle\psi|\,\hat{Q}\,|\psi\rangle = 1$: i.e. the elements of W to which the quantum formalism assigns a probability of 1.

Next, note that if $\langle\psi|\,\hat{Q}\,|\psi\rangle = 1$ then this is true for any coarse-graining of $\hat{Q}$, i.e. $\langle\psi|\,\hat{Q}\,|\psi\rangle = 1$ implies $\langle\psi|\,\hat{Q}'\,|\psi\rangle = 1$ for all $\hat{Q}'$ such that $\hat{Q} \preceq \hat{Q}'$. This means that the collection of subsets T^{ψ}_{W} of G_W, W in $\mathcal{W}$, forms a *sub-presheaf*, T^{ψ}, of G. Therefore, there is an associated characteristic arrow, defined in equation (19), and then the theory of presheafs says that the associated (contextual) truth values are sieves and hence belong to a Heyting algebra.

Rewriting equation (19) for the special case of the sub-object T^{ψ} of G we finally arrive at the following definition of a generalized valuation associated with a quantum state $|\psi\rangle$[21]. The generalized truth value of a projector $\hat{P}$, in the context of the window W to which $\hat{P}$ belongs, is the sieve on W defined by

$$\chi^{\psi}_{W}(\hat{P}) := \{W' \mid W' \subset W \text{ and } \langle\psi|\,G_{WW'}(\hat{P})\,|\psi\rangle = 1\}. \tag{22}$$

Thus the truth value, $\chi^{\psi}_{W}(\hat{P})$, associated with a projector $\hat{P}$ is (i) contextual: it depends on the window to we which we think of $\hat{P}$ as belonging; and (ii) multi-valued: $\chi^{\psi}_{W}(\hat{P})$ is a sieve on W and these form a Heyting algebra.

Note that if the state $|\psi\rangle$ is an eigenvector of $\hat{A}$ with an eigenvalue that lies in Δ, then $\langle\psi|\,\hat{E}[A \in \Delta]\,|\psi\rangle = 1$, and equation (22) gives the generalized truth value to

be the set of *all* coarse-grainings of W. This is knows as the *principal* sieve on W, and is the unit element of the Heyting algebra of sieves on W[22].

En passant, we remark that the discussion in Section 2.4 about macro-states in classical physics can be re-expressed in presheaf language, thereby providing a proof that the truth values are sieves, and hence form a Heyting algebra. The reader is referred to the original papers for further details on these, and related, matters [4–6].

4 CONCLUSION

We started by showing how, in classical physics, the idea of coarse-graining can be used to associate a generalized truth value with the proposition '$A \in \Delta$' for a macrostate M even if there are some states s in M for which $\overline{A}(s)$ does not belong to Δ.

Motivated by these ideas we turned to quantum theory, with the aim of using presheaf theory as a natural mathematical framework in which to discuss contextual, multi-valued, truth values. As the space of contexts we chose the set $\mathcal{W}$ of all Boolean subalgebras of the non-distributive logic of all projection operators. We then constructed two natural presheafs over this space of windows: the presheaf of local truth values and the coarse-graining presheaf G. For each quantum state $|\psi\rangle$ we constructed a special sub-object, T^{ψ} of G, and used this to define the quantum valuation in equation (22).

In short, we have shown that, notwithstanding the Kochen–Specker theorem, it *is* possible to assign truth-values to the projection operators in a quantum theory, but these truth values are both contextual and multi-valued. It is important to emphasise that the logical connectives ('and', 'or', 'not') are *uniquely* specified by the mathematics of topos theory as applied to presheafs.

However, in addition to being contextual, the presheaf logic differs from a simple Boolean algebra in that it is a *Heyting algebra*, and the principle of excluded middle, $P \vee \neg P = 1$, no longer holds. Equivalently, although it is still true that P implies $\neg\neg P$ it is no longer the case that $\neg\neg P$ implies P. In particular, this means that proofs by contradiction are no longer valid. This is a characteristic feature of, so-called, 'intuitionistic' logic, and (unlike non-distributivity) is easy to live with once one has got used to it.

Does all this mean that, after all, quantum theory *can* be interpreted in a realist way? Clearly the answer is 'no', if 'realist' is understood in the sense used in the Introduction—i.e. propositions about the world are handled using standard Boolean logic. For our truth values are contextual and multi-valued. On the other hand, the presheaf logic *is* distributive (unlike quantum logic proper) and *can* therefore be used as the basis for a deductive system for reasoning about the world. In this sense, our generalized truth values are closer to classical logic than quantum logic. Jeremy Butterfield and I have referred to the corresponding philosophical position as 'neo-realism'.

At this point, any physicist reader who has courageously slogged through the paper might well say 'Well done lads, but is it useful?'—a justified, but frequently embarrassing, question that is routinely addressed to any one claiming to have arrived

at a new result in the foundations of quantum theory. One response might be 'Well: our way of looking at things gives a better picture, or 'tells a better story', of what quantum theory is saying about the world. And this is valid in its own right'.

Personally, I think that this is true (but then I would, wouldn't I?) but nevertheless it would be good to be able to put the scheme to work in some concrete way. The obvious subject area is quantum cosmology, particularly cosmogenesis where the scheme could be used to handle statements about 'how thing are' in that very extreme stage of the universe. In this context it is worth remarking that our scheme can be viewed as a type of 'many-worlds' interpretation of quantum theory, with a 'world' being understood as a 'window on reality': i.e. a Boolean subalgebra of the non-distributive logic of all projectors. The actual working through of this structure in the context of a specific quantum cosmological model remains high on my list of research topics.

AUTHOR'S BIOGRAPHY

Chris Isham gained his PhD at Imperial College in 1969 working under the supervision of Paul Matthews. He then spent a year with Abdus Salam at the International Center for Theoretical Physics in Trieste, Italy. In 1970 he started his Lectureship in the Theoretical Physics group at Imperial College. In 1973 he moved to King's College, London, as a Reader in Mathematics, and returned to Imperial College in 1976. He was appointed to a Chair in 1982.

Chris Isham's main research interests are quantum gravity and mathematical aspects of the foundations of quantum theory. He also has a deep interest in general philosophy and the work of C.G. Jung.

NOTES

[1] Actually, the theorem only holds if the dimension of the vector space of states is greater than two, whereas in the example under discussion the dimension is equal to two. However, this does not alter the general thrust of the argument being developed here.

[2] A *valuation* is a function that assigns a value (a real number) to each physical quantity. When applied to propositions, a valuation assigns a truth value: 1 for 'true', and 0 for 'false'.

[3] Recall that a projection operator is a Hermitian operator $\hat{P}$ that satisfies $\hat{P}^2 = \hat{P}$.

[4] This is a rather loose way of speaking: the subspace $\mathcal{H}_{\hat{P}} \cap \mathcal{H}_{\hat{Q}}$ really represents *all* of the propositions '*P and Q*' as P and Q range over the propositions represented by $\hat{P}$ and $\hat{Q}$ respectively. Similar remarks apply to the logical 'or' and 'not' operations.

[5] The orthogonal complement of a subspace, W, of $\mathcal{H}$ is the set of all vectors that are orthogonal to every vector in W.

[6] In statistical physics, the macro-state M would be given some probability by the theory. Of course, an assignment of probabilities to macro-states is not incompatible with a realist view in which the system *has* a definite state (and each physical quantity has a definite value) but we happen not to know what this state is, only that it lies in the subset M of $\mathcal{S}$.

[7] Although one might want to handle the situations in which M is 'only just' a subset of $\mathcal{S}_{A\in\Delta}$, so that '$A \in \Delta$' is 'almost not totally true'.

[8] Here, $f(\Delta)$ denotes the set of all real numbers of the form $f(s)$ where s belongs to Δ.

[9] The plural 'measurements' arises in the relative frequency interpretation of probability. This is the interpretation normally used in instrumentalist approaches to quantum theory.

[10] If $\mathrm{Prob}(A\Delta;\ |\psi\rangle) = 1$, might we also want to consider the possibility that $|\psi\rangle$ is 'close' to a state for which the probability is less than 1?

[11] Of course, something similar happens in classical physics, where many different propositions are represented by the same subset of the state space $\mathcal{S}$. However, the singular feature of quantum theory is that equation (9) can hold even if $[\,\hat{A}, \hat{B}\,] \neq 0$. The equality in equation (9) means that $\hat{A}$ and $\hat{B}$ have some simultaneous eigenvectors, but they are not a complete set if $[\,\hat{A}, \hat{B}\,] \neq 0$.

[12] Strictly speaking, only *Borel* subsets should be considered, but we will ignore such niceties.

[13] In general, $\hat{E}[f(A) \in J\,] = \hat{E}[\,A \in f^{-1}(J)\,]$ for any Hermitian operator $\hat{A}$ and (Borel) subset J of real numbers. Here $f^{-1}(J)$ denotes the set of all real numbers s such that $f(s)$ belongs to J.

[14] Equivalently, W_A contains the projectors onto the *simultaneous* eigenstates of $\hat{H}$ and $\hat{L}_x$, and ditto for W_B with $\hat{H}$ and $\hat{L}_y$.

[15] The notation $W_1 < W_2$ includes the possibility $W_1 = W_2$.

[16] This means that (i) for all W we have $W < W$; (ii) $W_1 < W_2$ and $W_2 < W_1$ implies $W_1 = W_2$; and (iii) $W_1 < W_2$ and $W_2 < W_3$ implies $W_1 < W_3$.

[17] The various spaces introduced at this point are all required to be topological spaces. In most applications in theoretical physics they are also differentiable manifolds.

[18] For a product bundle, there is a one-to-one correspondence between cross-sections and maps from B to the fibre F. A cross-section is then analogous to the, so-called, *graph* of a function as discussed in elementary, school-level mathematics.

[19] Actually, this is a very special type of presheaf. In general, a presheaf is defined over a *category*, and a poset is a particularly simple example of a category.

[20] It is also required that, for each p in $\mathcal{P}$, the map X_{pp} is the identity map from X_p to itself.

[21] These ideas can be trivially extended to a density matrix state, $\hat{\rho}$, using the fact that $\mathrm{Prob}(P; \hat{\rho}) = \mathrm{tr}(\hat{\rho}\hat{P})$ is the probability associated with the proposition P in the state $\hat{\rho}$.

[22] The null element is the empty set of sieves.

REFERENCES

[1] S. Kochen and E.P. Specker, The problem of hidden variables in quantum mechanics. *J. Math. Mech.* 17 59 (1967).

[2] H. Reichenbach. *Philosophic Foundations of Quantum Mechanics* (Dover Publications, New York, 1998).

[3] C.J. Isham, *Lectures on Quantum Theory: Mathematical and Structural Foundations* (Imperial College Press, London, 1995).

[4] C.J. Isham and J. Butterfield, A topos perspective on the Kochen-Specker theorem: I. Quantum states as generalized valuations. *Int. J. Theor. Phys.* **37** 2669 (1998).

[5] J. Butterfield and C.J. Isham, A topos perspective on the Kochen-Specker theorem: II. Conceptual aspects, and classical analogues. *Int. J. Theor. Phys.* **38** 827 (1999).

[6] J. Hamilton, J. Butterfield and C.J. Isham, A topos perspective on the Kochen-Specker theorem: III. Von Neumann algebras as the base category. *Int. J. Theor. Phys.* **39** 1413 (2000).

[7] R. Goldblatt, *Topoi: The Categorial Analysis of Logic*, (North-Holland, London 1984).

HERBERT KORTÉ*

9. EINSTEIN'S HOLE ARGUMENT AND WEYL'S FIELD-BODY RELATIONALISM†

ABSTRACT

Einstein's hole problem concerns the nature of causality in the General Theory of Relativity (GTR). This paper introduces a number of formal concepts and distinctions which are necessary for a clear understanding of the nature of causality in GTR. In particular, the distinctions between formal, theoretic and physical coordinates are introduced as well as the distinction between model and symmetry transformations. Utilizing the notion of local diffeomorphisms which are globally defined locally invertible maps, it is made explicit that model diffeomorphisms and passive coordinate transformations are mathematically equivalent. This, it is argued, decisively undercuts the claims by Earman and Norton that a spacetime substantivalist view is faced with 'radical local indeterminism' for a range of modern spacetime theories, including GTR. Additional epistemic and ontological difficulties in Earman's and Norton's accounts of Einstein's hole argument are discussed, and it is argued that these difficulties underscore the need to adopt the ontological position called 'field-body relationalism', a position forcefully advanced by Hermann Weyl. The paper concludes with a discussion of Weyl's critique of body-relationalism, Weyl's argument for the necessity of the inertial field (guiding field), and a modern re-formulation of Newton's laws of motion that explicitly takes account of Weyl's field-body-relationalist ontology.

1 INTRODUCTION

Several years prior to the completion of the final version of GTR, Einstein was very much concerned with the question whether a law of gravitation that satisfies the principle of general covariance can also satisfy the principle of causality. At that time Einstein arrived by means of his hole argument at the conviction that the spacetime metric field equations could not be generally covariant *and* also satisfy the principle of causality.[1] Einstein's final version of the hole argument is as follows. Consider a solution $g_{ij}(x)$ with respect to a given coordinate system x for given initial conditions on an initial surface S. Suppose that there is a matter-free region Σ (a hole) that is future-related to S. Apply an active diffeomorphism which is the identity outside the region Σ. Then the dragged-along metric tensor $\tilde{g}_{ij}(x)$ is also a solution of the field equations and differs from $g_{ij}(x)$ only inside Σ. Thus with respect

* Department of Philosophy, University of Regina.

† This paper owes a great deal to earlier collaboration with Robert Coleman. Robert died June 30, 2001.

W. Demopoulos and I. Pitowsky (eds.), Physical Theory and its Interpretation, 183–212.

to the same coordinate system, one has two distinct solutions of the field equations both of which satisfy the same initial value conditions. Einstein repeatedly returned to this problem and later concluded on the basis of another argument (his 'coincidence argument')[2] that general covariance is compatible with physical determinism after all.

Einstein's hole argument is strikingly similar in content to the modern formulation of the Cauchy or Initial Value problem, which may briefly be characterized as follows. Assuming that spacetime $\langle M, g\rangle$ is globally hyperbolic, there exists a one parameter foliation of spacetime the leaves of which are Cauchy hypersurfaces. For a given choice of an initial Cauchy hypersurface S, the appropriate initial data consists of the Riemannian metric h induced on S by the spacetime metric g and the extrinsic curvature tensor K for the embedding of $\langle S, h\rangle$ in $\langle M, g\rangle$. If g satisfies the Einstein field equations, then the tensors h and K are not independent but satisfy certain constraint equations. Moreover, the spacetime $\langle M, g\rangle$ is determined by $\langle S, h, K\rangle$ up to an isometry; that is, if $\langle M, \tilde{g}\rangle$ is any other spacetime for which S is a Cauchy surface, h is obtained by restricting $\tilde{g}$ to S and K is the extrinsic curvature tensor for the embedding of $\langle S, h\rangle$ in $\langle M, \tilde{g}\rangle$, then there exists a diffeomorphism $f : M \to M$ such that $g = f^*\tilde{g}$. If material bodies are included, whether sources or test-bodies, the initial data includes the points at which the world line of each body pierces S and the vector that represents the spatial component of the unit four velocity of the body at that point. Although the presence of sources greatly increases the mathematical difficulties, the spacetime and the world lines of the material bodies are again determined up to a diffeomorphism.

One important feature of the Cauchy problem for GTR is that the initial data determines the solution of Einstein's field equations only up to a diffeomorphism. What implication does this feature of the Cauchy problem have for the principle of causation and for the measurement of the spacetime metric? *Prima facie* the following claims are incompatible:

1. The spacetime metric coefficients $g_{ij}(x^i)$ can be uniquely determined empirically.
2. The solutions to Einstein's equations are unique only up to an active diffeomorphism, given an initial data set on a portion of a Cauchy hypersurface.

There is general agreement that GTR satisfies the requirements of physical causality. However, the reasons advanced for holding this view are varied and far from coherent. In the physics literature, the fact that the solution in the domain of dependency of the initial data is determined *only* up to a local diffeomorphism in any local neighbourhood of this domain is explained away in a variety of ways: diffeomorphically equivalent models are asserted to be physically equivalent; the active diffeomorphism is asserted to be equivalent to a passive coordinate transformation and hence the lack of uniqueness comes down to the necessity of making a coordinate choice; the lack of uniqueness is similar to the need for a choice of gauge in electromagnetism.[3] While these explanations are not false in any straight-forward sense, they are unsatisfactory for a number of reasons. Moreover, Einstein's coincidence argument is sometimes advanced as the basis for asserting the physical equivalence of diffeomorphically equivalent models. However, the set of coincidences of material bodies would not

suffice to establish physical equivalence because the data base would not be sufficiently rich. It is also not clear how an active diffeomorphism can be equivalent to a passive transformation and yet be intimately associated with symmetry transformations. Moreover, to remark that the manifold diffeomorphism group is the gauge group of GTR is nothing but an appeal to a loose analogy.

The organization of the paper is as follows: In the first part of the paper, additional concepts and distinctions are introduced which are necessary for a clear understanding of the nature of causality in GTR. In particular, we emphasize the necessity of distinguishing between formal, theoretic, and physical coordinates, as well as between *model* and *symmetry* diffeomorphisms. We clarify the relationship between model diffeomorphisms and passive coordinate transformations, and discuss the importance and physical significance of a number of other types of transformations that are relevant to the Initial Value problem in GTR. In particular, utilizing the notion of *local diffeomorphisms* which are globally defined locally invertible maps, we make explicit that model diffeomorphisms and passive coordinate transformations, though conceptually quite different, are mathematically equivalent. This undercuts the arguments of Earman (1986, 1989), Norton (1987, 1988, 1989, 1992) and Earman and Norton (1987) which attempt to establish that a spacetime substantivalist view is faced with 'radical local indeterminism' for a range of modern spacetime theories, including GTR. For it shows that their sense of an *active* transformation in the context of their discussion of Leibniz equivalence of spacetime models is mathematically equivalent to a passive reading. Moreover, additional difficulties of an epistemological and ontological nature in Earman's and Norton's accounts underscore the need to adopt what we call 'field-body relationalism', an ontological position forcefully advanced by Hermann Weyl. Though compatible with a spacetime substantivalist view, Weyl's position strongly suggests that the spacetime manifold M should be viewed merely as a conceptual scaffolding, a mental construct necessary only at the initial stages of the modeling process. The paper concludes with a brief discussion of Weyl's critique of body-relationalism championed by Leibniz, Mach and Einstein, Weyl's argument for the necessity of the inertial field (guiding field) and a modern re-formulation of Newton's laws of motion that explicitly takes account of Weyl's field-body relationalist ontology.

2 FORMAL, EPISTEMOLOGICAL AND ONTOLOGICAL CONSIDERATIONS

The term 'active' is usually construed to include *both* model and symmetry diffeomorphisms. But this is very misleading for the following reasons: When one speaks of a model-diffeomorphism one has in mind an *active* transformation such that the bodies and fields are *actually* moved either physically or conceptually; that is, the diffeomorphisms are in some sense 'actually carried out' or 'executed'. In contrast, a symmetry transformation does not involve any such *actual* movement or execution at all, but instead involves a *comparison* of a computed image of some portion of the world with the original state. Therefore, the notion of 'diffeomorphism' used in the

Table 9.1 Typology of transformations

		Basic types of transformations		
		Active	*Passive*	*Symmetry*
Coordinate type	*Formal*	Formal-active	Formal-passive	Formal-symmetry
	Theoretic	Theoretic-active	Theoretic-passive	Theoretic-symmetry
	Physical	Physical-active	Physical-passive	Physical-symmetry

context of 'symmetry' must clearly be distinct from the one used in the context of 'diffeomorphically equivalent models'. However, the language currently used in the literature does not support the distinction between these two radically different concepts of a diffeomorphism. To bring out these distinctions explicitly we shall add to the basic *active* and *passive* transformations a third, namely *symmetry* transformations.

2.1 Types of Coordinate Systems

There are three, qualitatively different kinds of coordinate systems that are required for the construction of an adequate model of the physical world. These types of coordinates, will be distinguished by the labels *formal*, *theoretic* and *physical*. The basic active, passive and symmetry transformations are each subdivided according to the kind (formal, theoretic or physical) of coordinate system(s) used to describe it. Table 9.1 summarizes the results.

Remark 2.1 There are seven kinds of passive transformations that are of interest: the three simple passive transformations listed in table 9.1 and four *mixed*-passive transformations. We will briefly describe some of the *mixed*-passive transformations in remark 2.4.

Formal coordinates are purely abstract, mathematical coordinates that are used by the theorist to model the contents of the world, the dynamics and interactions of the various physical entities of the world and the physical procedures used to survey the world. Typically, it is stated that an n-dimensional, differentiable manifold M is a Hausdorff, topological space equipped with an atlas, that is, a family $\{(U_\alpha, x_\alpha)\}$ of charts, such that the open neighbourhoods U_α cover M. The maps[4] $x_\alpha : U_\alpha \to x_{\alpha\vdash}(U_\alpha) \subseteq \mathbb{R}^n$ are homeomorphisms, and whenever $U_\alpha \cap U_\beta \neq \emptyset$, the coordinate transformation maps

$$x_\beta \circ x_\alpha^{-1} : x_{\alpha\vdash}(U_\alpha \cap U_\beta) \to x_{\beta\vdash}(U_\alpha \cap U_\beta)$$

are C^k for $1 < k < \infty$, C^∞ or C^ω. The coordinate charts (U_α, x_α) are formal.

Even in purely mathematical accounts of certain geometric structures, special coordinate systems, called *adapted* coordinate systems, are introduced that are determined by the geometric structure in question. Such coordinate systems provide an *internal* description of the particular geometric structure, a description that is more

faithful in various respects. We call such coordinates *theoretic* coordinates because theoretical coordinates are determined by some fundamental entity, typically the geometric structure of the world, that is postulated by the theory. Whereas a formal coordinate description is completely arbitrary, a priori and 'external-to-the-structure', a theoretic coordinate description, by contrast, is 'internal-to-the-structure' and is, in the case of geometry, determined possibly up to a Lie group of symmetry transformations of the geometry. For example, in Galilean models of spacetime, the spatial geometry is typically introduced by stipulating the existence of a system of spatial coordinates (x, y, z) with respect to which the metric is given by $ds^2 = dx^2 + dy^2 + dz^2$. These coordinates are determined up to a Euclidean transformation by a fundamental element of the theory, the spatial metric, and are hence linked to the metrical structure postulated by the theory. Similarly in the Special Theory of Relativity, it is assumed that a system of coordinates (t, x, y, z) exists with respect to which the spacetime metric is given by $ds^2 = -dt^2 + d\vec{r} \cdot d\vec{r}$. These coordinates are likewise *theoretic* and are determined up to a Poincaré transformation by the fact that they are adapted to a fundamental element of the theory, the spacetime metric.

A *physical* coordinate system is one that makes use of various physical entities, bodies and fields, to assign coordinates to physical events. It is this kind of coordinate system that is used by a *physical observer* to track material bodies and to measure various physical fields such as the electromagnetic field and the spacetime-metric field. An example of a physical coordinate chart is the radar tracking system that may be found at every major airport. A physical atlas for a region surrounding the earth is provided by the Global Positioning System.

2.2 Active transformations

2.2.1 Physical-active transformations The physical-active transformation is the classical case that underlies our intuitions regarding diffeomorphisms. It is this type of transformation that arises in the mechanics of continua when an elastic or a fluid body is subjected to forces which cause it to deform smoothly from an initial state to a final state. In general, the initial state is destroyed in the process of creating the final state.

It should be noted that a physical-active transformation can actually be executed only because material of limited spatial extent is acted on and there exists an additional dimension, time, through which the transformation can be effected. In addition, a physical-active transformation has a straight-forward interpretation only if the spacetime curvature in the region considered is small with respect to the spatial extension and temporal duration of the physical-active transformation. In such a region, one can set up a system of physical coordinates that are macroscopically adapted to the essentially flat local spacetime geometry. With respect to such a system of coordinates, the physical-active transformation can be characterized as a one parameter family of local diffeomorphisms, a flow in the spatial region under consideration.

2.2.2 Formal-active transformations The formal-active transformation occurs in discussions about diffeomorphically equivalent models of spacetime theories;

moreover, it is this sense of the term 'diffeomorphism' that is employed in statements asserting that the solution of the Initial Value problem in GTR is unique up to a diffeomorphism. In terms of the modeling process, a formal-active transformation re-locates the *entire* physical content of the world—everything, the spacetime metric, electromagnetic and other physical fields as well as material bodies—in its 'container'[5], the spacetime manifold M. Since everything that is needed for the specification of either physical coordinates or theoretic coordinates is affected by a formal-active transformation, it is clear that such a transformation can only be described with respect to an atlas of coordinate charts that are purely *formal*.

There is some similarity here with the indiscernible transformation considered by Leibniz who considered re-locating the entire physical content of the world. If one allows for the existence of a homogeneous, absolute metric of space, then such a re-location of the physical content other than the metric, constitutes a Euclidean motion which can be described with respect to a theoretic coordinate system adapted to the spatial Euclidean metric. The formal-active and theoretic-active transformations are then both physically indiscernible. The re-location associated with a formal (theoretic)-active transformation occurs only in the mind of the theorist who now plays the role assigned to God by Leibniz.

2.2.3 Symmetry transformations Although the formal, theoretic and physical viewpoints significantly modify the content and description of a symmetry transformation, the existence or non-existence of a symmetry is coordinate independent. The following characterization applies equally to all three coordinate types of symmetry-transformations.

One has a description of the world (or of a portion of the world or a model thereof) with respect to a *fixed* atlas, and considers a diffeomorphism f described for example as in proposition 2.2 below. An *image* of what the physical content of the world *would* be if the physical fields and material bodies *were* dragged along by the diffeomorphism is *computed*; that is, for each μ, under the image provided by $f_\mu\colon U_\mu \to V_\mu$, the contents of the region U_μ are mapped onto the region V_μ. The diffeomorphism f is a symmetry of some physical entity if and only if for every μ, the descriptors of that physical entity in V_μ are the same as the descriptors of the image under f_μ of the corresponding physical entity in U_μ, where the term 'descriptors' denotes the set of components that describe the physical entity with respect to the relevant coordinate chart.

In contrast with a formal-active transformation, the image under a symmetry transformation is computed *as if* it were carried out, but it is *not* regarded as actually carried out. The counterfactual, pseudo, simulated or computed nature of symmetry transformations is not really made explicit in the literature although some presentations (e.g. Wald, 1984, 438) employ scare quotes. Another notable contrast is that a *comparison* is made in the case of a symmetry transformation. On the other hand, no particular agreement is expected to exist between the initial and final states related by a formal-active transformation. In addition, a formal-active transformation always applies to *every* physical entity; however, a symmetry transformation may be applied *selectively*.

One could, for example, look for the symmetries of just the electromagnetic field or of just some aspect of the geometry such as the projective or conformal structure.

2.2.4 Representation of Model and Symmetry Diffeomorphisms The following proposition about the representation of diffeomorphisms will be used below. The proposition applies to both symmetry and model diffeomorphisms.

Proposition 2.2 *Given any atlas for M and any diffeomorphism $f: M \to M$, one can construct atlases $\{(U_\mu, x_\mu)\}$ and $\{(V_\mu, y_\mu)\}$ such that for any μ,*

$$f_{\vdash}(U_\mu) = V_\mu,$$

where f is determined by local diffeomorphisms $f_\mu: U_\mu \to V_\mu$ which satisfy

$$f_\mu|_{U_\mu \cap U_\nu} = f_\nu|_{U_\mu \cap U_\nu}$$

whenever $U_\mu \cap U_\nu \neq \emptyset$; moreover, the f_μ are represented in $\mathbb{R}^n$ by the local diffeomorphisms $F_\mu = y_\mu \circ f_\mu \circ x_\mu^{-1}$.

Proof. Let $f: M \to M$ be an arbitrary diffeomorphism and $\{(B_\alpha, s_\alpha)\}$ be an arbitrary atlas. For each α, define the open sets A_α and C_α by

$$A_\alpha = f^{\dashv}(B_\alpha) \quad \text{and} \quad f_{\vdash}(B_\alpha) = C_\alpha.$$

Then

$$f_{\vdash}(A_\alpha \cap B_\beta) = f_{\vdash}(A_\alpha) \cap f_{\vdash}(B_\beta) = B_\alpha \cap C_\beta.$$

Since

$$\bigcup_{\alpha\beta} A_\alpha \cap B_\beta = M = \bigcup_{\alpha\beta} B_\alpha \cap C_\beta,$$

both of the collections $\{A_\alpha \cap B_\beta\}$ and $\{B_\alpha \cap C_\beta\}$ are open covers of M; consequently, $f: M \to M$ is determined by a collection of local diffeomorphisms

$$f_{\alpha\beta}: A_\alpha \cap B_\beta \to B_\alpha \cap C_\beta$$

that satisfy

$$f_{\alpha\beta}|_{(A_\alpha \cap B_\beta) \cap (A_\gamma \cap B_\delta)} = f_{\gamma\delta}|_{(A_\alpha \cap B_\beta) \cap (A_\gamma \cap B_\delta)}$$

whenever $(A_\alpha \cap B_\beta) \cap (A_\gamma \cap B_\delta) \neq \emptyset$.

Let $s_\alpha: B_\alpha \to S_\alpha \subseteq \mathbb{R}^n$ be the coordinate maps of the original atlas. Define the coordinate map $r_{\alpha\beta}: A_\alpha \cap B_\beta \to R_{\alpha\beta}$ by restricting the map s_β and define the coordinate map $t_{\alpha\beta}: B_\alpha \cap C_\beta \to T_{\alpha\beta}$ by restricting the map s_α. Then each of the

collections $\{(A_\alpha \cap B_\beta, r_{\alpha\beta})\}$ and $\{(B_\alpha \cap C_\beta, t_{\alpha\beta})\}$ is an atlas for M. The functions $f_{\alpha\beta}$ are represented by the functions

$$F_{\alpha\beta} = t_{\alpha\beta} \circ f_{\alpha\beta} \circ r_{\alpha\beta}^{-1} : R_{\alpha\beta} \to T_{\alpha\beta}.$$

The result is obtained in the form stated in the proposition by making the following substitutions: $(A_\alpha \cap B_\beta, r_{\alpha\beta}) \to (U_\mu, x_\mu)$, $(B_\alpha \cap C_\beta, t_{\alpha\beta}) \to (V_\mu, y_\mu)$, $f_{\alpha\beta} \to f_\mu$ and $F_{\alpha\beta} \to F_\mu$. □

2.3 *Passive transformations*

From a *purely mathematical* point of view, all passive transformations (formal, physical, theoretic) are described in the same way. Let (U, x) and $(\bar{U}, \bar{x})$ be two coordinate charts belonging to an atlas for M, and suppose that $\bar{U} \cap U \neq \emptyset$. The coordinate transformation from (U, x) to $(\bar{U}, \bar{x})$ is determined by the local diffeomorphism of the *coordinate space* given by

$$\bar{X}^i = \bar{x}^i \circ x^{-1} : x_{\vdash}(\bar{U} \cap U) \to \bar{x}_{\vdash}(\bar{U} \cap U). \tag{1}$$

The inverse of this map is denoted by $X^i = x^i \circ \bar{x}^{-1}$.

A geometric object at a point $p \in M$ is determined by its components with respect to a local coordinate chart. If $p \in \bar{U} \cap U$, the geometric object has a set of components or descriptors with respect to each of the charts $(\bar{U}, \bar{x})$ and (U, x). The map $\bar{X}^i(x^i)$ and its inverse $X^i(\bar{x}^i)$ determine the functional relations that express each set of components in terms of the other set. Consider, for example, a pseudo-Riemannian metric represented by

$$\bar{g}(p) = \bar{g}_{ij}(\bar{x}^i(p)) d_p\bar{x}^i \otimes d_p\bar{x}^j$$

in $\bar{U}$ and by

$$g(p) = g_{ij}(x^i(p)) d_p x^i \otimes d_p x^j$$

in U.

If $p \in \bar{U} \cap U$, these forms must be identical, and since

$$d_p x^i = X^i_j(\bar{x}^i(p)) d_p\bar{x}^j,$$

where $X^i_j(\bar{x}^i)$ is the partial derivative of $X^i(\bar{x}^i)$ at $\bar{x}^i(p)$, the transformation law is given by

$$\bar{g}_{ij}(\bar{x}^i(p)) = g_{rs}(X^i(\bar{x}^i(p))) X^r_i(\bar{x}^i(p)) X^s_j(\bar{x}^i(p)). \tag{2}$$

If both of the coordinate charts (U, x) and $(\bar{U}, \bar{x})$ are formal, then (1) is a formal-passive transformation and the transformation law (2) defines a constraint on the otherwise free choice of the metric component functions. On the other hand,

if both charts are physical, then (1) is a physical-passive transformation and the law has empirical content insofar as Ehlers et al. (1972) Constructive Axiomatics supplemented by Coleman and Korté's (1980, 1982, 1987, 1989, 1990, 1991, 1992ab) measurement procedure for directing fields provides a noncircular procedure for the unique, empirical determination of the metric coefficients $g_{ij}(x^i(p))$ and $\bar{g}_{ij}(\bar{x}^i(p))$. Moreover, it is possible to carry out measurements to determine empirically the transformation functions $X^i(\bar{x}^i(p))$, albeit with great difficulty in the general case. Since the various factors in the transformation law (2) can be empirically determined independently, the transformation law (2) can be used to verify that the empirically determined coefficients $g_{ij}(x^i(p))$ and $\bar{g}_{ij}(\bar{x}^i(p))$ actually do satisfy the transformation law for the coefficients of a second order covariant tensor field.

The following proposition provides a slightly more general description of passive transformations. Such a generalized description is needed both to explain the notion of a *mixed*-passive transformation and to show that each formal-active transformation (model diffeomorphism) corresponds to a formal-passive transformation. The proposition provides a standard representation of a passive transformation for the case in which *two* entire atlases are involved.

Proposition 2.3 *Let* $\{(A_\alpha, r_\alpha)\}$ *and* $\{(B_\beta, s_\beta)\}$ *be two atlases for* M *which determine the same differentiable structure on* M. *Then, there exists a refinement* $\{(U_\mu, x_\mu)\}$ *of* $\{(A_\alpha, r_\alpha)\}$ *and a refinement* $\{(U_\mu, z_\mu)\}$ *of* $\{(B_\beta, s_\beta)\}$ *such that the maps*

$$Z^i_\mu = z^i_\mu \circ x^{-1}_\mu : x_{\mu\vdash}(U_\mu) \to z_{\mu\vdash}(U_\mu)$$

and their inverses $X^i_\mu = x^i_\mu \circ z^{-1}_\mu$ *determine the transformation from* $\{(A_\alpha, r_\alpha)\}$ *to* $\{(B_\beta, s_\beta)\}$.

Proof. Define the enrichment of $\{(A_\alpha, r_\alpha)\}$ to be $\{(A_\alpha \cap B_\beta, r_{\alpha\beta})\}$, where $r_{\alpha\beta}$ is the restriction of r_α. Similarly, the enrichment of $\{(B_\beta, s_\beta)\}$ is defined to be $\{(A_\alpha \cap B_\beta, s_{\alpha\beta})\}$, where $s_{\alpha\beta}$ is the restriction of s_β. Now, re-label the doubly indexed quantities; so that, U_μ is $A_\alpha \cap B_\beta$, x_μ is $r_{\alpha\beta}$ and z_μ is $s_{\alpha\beta}$. □

Remark 2.4 Of course, this generalized description of a passive transformation may be used for the three cases in which both atlases are formal, theoretic or physical. Rather more interesting, however, are two cases of *mixed*-passive transformations. Consider a coordinate transformation between a description of the world with respect to a formal system of coordinates and one with respect to a physical system of coordinates. This type of mixed-passive transformation is a necessary component of the ADM-formulation[6] of the physical Cauchy problem for GTR which is discussed in Coleman and Korté (1992b). Briefly, after carrying out a survey of a portion of spacetime and computing the physical initial data on a chosen Cauchy surface, one abandons the physical radar-station coordinate systems used in the surveying process in favour of a system of *formal* coordinates that are defined by stipulating the so called *lapse* and *shift* functions. The solution obtained provides a purely formal image of the world; however, this image *includes* the motions of the bases of the original physical

radar-station coordinate systems along with the motions of other bodies and fields with the consequence that one can *compute* from the formal description *precisely* what any one of the physical observers will observe in the future domain of dependency of the Cauchy surface chosen in the originally surveyed region of spacetime. In particular, one can compute what physical coordinates a given physical observer would assign to a point designated by a particular set of formal coordinates; that is, one can compute the *physical-formal*-passive transformation.

Hilbert had another approach to the physical Cauchy problem. Instead of abandoning the physical radar coordinate charts in favour of a purely formal system of coordinates as in the ADM approach, one solves the equations of motion with respect to a *theoretic* system of coordinates determined in terms of the spacetime metric, by imposing, for example, coordinate conditions on the metric such as the conditions for harmonic coordinates. The solution then provides a theoretic description of the world and one must compute the relevant *physical-theoretic*-passive transformation in order to know what a given physical observer will observe.

2.4 The equivalence of formal-active and formal-passive transformations

In the process of working through the difficulties associated with his hole problem, Einstein came to realize that a formal-active transformation is equivalent to a formal-passive transformation at least in the case of a region that can be covered by a single coordinate chart[7].

For the case of a *single* coordinate neighbourhood it is easy to show that to each formal-active transformation there corresponds a formal-passive transformation and conversely. Consider a local diffeomorphism $f: U \to U$. Then in the most general case, one may consider two different coordinate charts (U,x) and (U,y) for the neighbourhood U and describe the local diffeomorphism f by the local diffeomorphism in $\mathbb{R}^n$

$$F = y \circ f \circ x^{-1} : x\vdash(U) \to y\vdash(U).$$

Clearly, one can define another local chart (U,z) where $z = y \circ f$. The components that describe a geometric entity in U with respect to (U,z) are the same as the components that describe the entity transformed by f with respect to the chart (U,y). The situation is illustrated in Figure 9.1.

On the other hand, consider a formal-passive transformation from the coordinate chart (U,x) to the coordinate chart (U,z) with the transition function $Z = z \circ x^{-1}$. To recover the previous active picture relative to the two charts (U,x) and (U,y) one need only set $f = y^{-1} \circ z : U \to U$. The situation is illustrated in Figure 9.2.

Note that if (U,y) is chosen to be the same as (U,x), then the relationship between the active and passive descriptions is unique up to the choice of the chart (U,x). There is, however, no necessity to use the same chart 'before' and 'after' the local diffeomorphism f. Moreover, there is no similar canonical choice that may be used

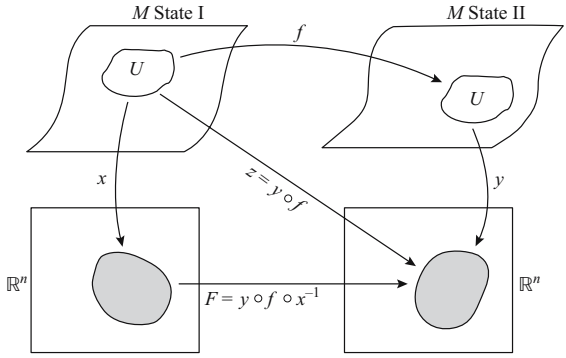

FIGURE 9.1. Simple active to passive.

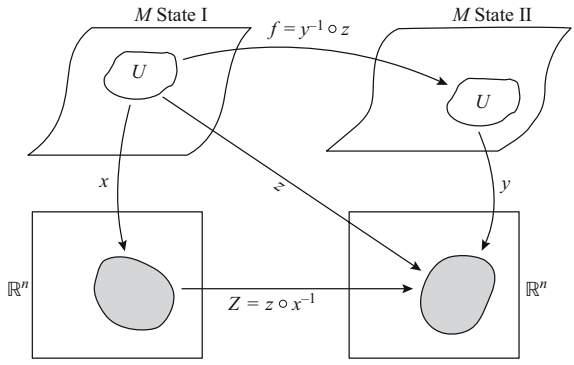

FIGURE 9.2. Simple passive to active.

to eliminate the corresponding freedom in the more general case which we shall now discuss.

The next proposition says that the equivalence of formal-active and formal-passive transformations holds even when such transformations are *globally* defined.

Proposition 2.5 (Equivalence of formal-active/passive) *Let $f : M \to M$ denote a formal-active transformation described with respect to some atlas for M, then there exists an infinite number of formal-passive transformations that are equivalent to the formal-active transformation f. Conversely, given the formal-passive transformation between any two atlases for M, there exists an infinite number of equivalent formal-active transformations.*

Proof. Let $f : M \to M$ be a formal-active transformation described with respect to some atlas. Then by proposition 2.3, there exist enrichments of this atlas, $\{(U_\mu, x_\mu)\}$ and $\{(V_\mu, y_\mu)\}$ such that f is described by local diffeomorphisms $f_\mu : U_\mu \to V_\mu$. Define an atlas $\{(U_\mu, z_\mu)\}$ by setting $z_\mu = y_\mu \circ f_\mu$. Then, the

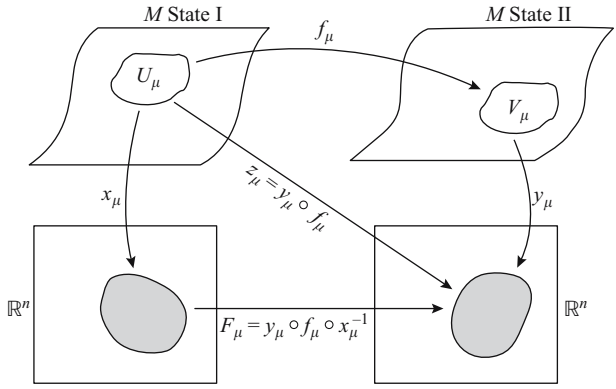

FIGURE 9.3. Active to passive.

formal-passive-transformation from the atlas $\{(U_\mu, x_\mu)\}$ to the atlas $\{(U_\mu, z_\mu)\}$ has the form stated in proposition 2.3 and is equivalent to f. Note that there exist an infinite number of formal-passive transformations that are equivalent to f because one gets a different one for each atlas with respect to which f may be described. Figure 9.3 illustrates the situation.

By proposition 2.3, given any two atlases for M, they may be refined to yield atlases $\{(U_\mu, x_\mu)\}$ and $\{(U_\mu, z_\mu)\}$, such that, the intra atlas transformation functions are given by $Z_\mu = z_\mu \circ x_\mu^{-1}$ and $X_\mu = x_\mu \circ z_\mu^{-1}$. Any diffeomorphism $f : M \to M$, however described, determines the image sets $V_\mu = f_\vdash(U_\mu)$ and the local diffeomorphisms $f_\mu : U_\mu \to V_\mu$. If one defines the atlas $\{(V_\mu, y_\mu)\}$ by setting $y_\mu = z_\mu \circ f_\mu^{-1}$, then the formal-active transformation f is described as in proposition 2.2 and is clearly equivalent to the given formal-passive transformation. Clearly, there are an infinite number of formal-active transformations that are equivalent to a given formal-passive transformation. Figure 9.4 illustrates the situation. □

In the case of GTR the metric is not an absolute structure with convenient global symmetries; rather, it is a dynamical physical entity that is coupled to the energy-momentum density of matter and other fields. In almost all cases, the spacetime metric does not have any symmetries at all[8]. The theory tells us that the best we can do is to adapt to a microneighbourhood by employing a normal coordinate system at a given point. In the absence of symmetry, extended bodies adapted to the geometry cannot exist. In such circumstances, neither theoretic nor physical coordinate systems can be introduced early in the presentation of the theory even with the aid of ad hoc assumptions, because the circumstances that permitted ad hoc assumptions in the pre-GTR case simply do not obtain in the context of GTR.[9]

A proper theoretical account of the epistemology of geometry requires an analysis of the physical measurement process itself, in particular an analysis of physical coordinate systems. In contrast with the geometric structures of classical mechanics, the geometric structure of GTR is not given a priori, rather, it is part of the dynamical

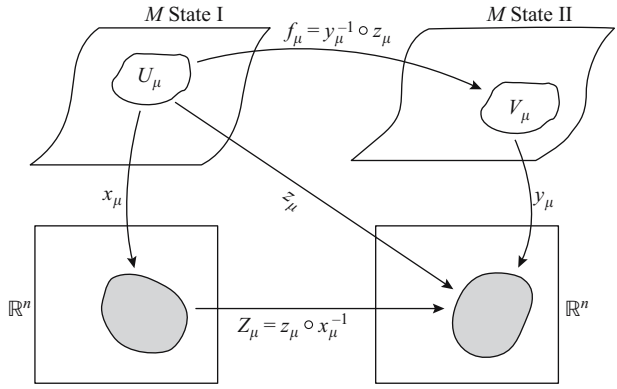

FIGURE 9.4. Passive to active.

problem, and hence theoretic and physical coordinates are also part of the dynamical problem.[10] In particular, the description of physical coordinate systems in GTR is really an advanced topic in the sense that many other physical entities, including the spacetime metric, the electromagnetic field, the world lines of material bodies and their equation-of-motion structures, have to be introduced and analyzed first. Clearly, the description and analysis of these physical entities must be carried out with respect to a purely mathematical or formal system of coordinates because theoretic and physical coordinates are not yet part of the model. In a little book entitled *Riemanns geometrische Ideen, ihre Auswirkung und ihre Verknüpfung mit der Gruppentheorie*, published posthumously in 1988, (Weyl, 1988, 4–5) makes this interesting comment:

> Coordinates are introduced on the Mf [manifold] in the most direct way through the mapping onto the number space, in such a way, that all coordinates, which arise through one-to-one continuous transformations, are equally possible. With this the coordinate concept breaks loose from all special constructions to which it was bound earlier in geometry. In the language of relativity this means: The coordinates are not measured, their values are not read off from real measuring rods which react in a definite way to physical fields and the metrical structure, rather they are a priori placed in the world arbitrarily, in order to characterize those physical fields including the metric structure numerically. The metric structure becomes through this, so to speak, freed from space; it becomes an existing field within the remaining structure-less space. Through this, space as form of appearance contrasts more clearly with its real content: The content is measured after the form is arbitrarily related to coordinates.

Weyl's statement that the "metric structure becomes through this, so to speak, freed from space; it becomes an existing *field* within the remaining structure-less space" and Weyl (1929) statement that "the metric field has been freed from the manifold" may both be understood mathematically in the following way: In his

Erlanger Programme, Felix Klein provided a unified approach to the various 'global' geometries by showing that each of the geometries is characterized by a particular group of transformations. As Weyl (1929) noted, it was E. Cartan who first adapted Klein's Erlanger Programme to infinitesimal geometry by applying Klein's notions to the tangent (or co-tangent) plane, rather than to the manifold itself and thereby founded the theory of $\mathfrak{G}$-structures. A first order $\mathfrak{G}$-structure, where $\mathfrak{G}$ is a subgroup of the general linear group $\mathrm{GL}(\mathbb{R}^n)$, is determined by specifying in a smooth manner an equivalence class of privileged frames or co-frames for the tangent or co-tangent spaces at every point of the manifold. At a given point p, any two equivalent frames or co-frames are related by an element of the group G in question which characterizes the infinitesimal geometry.

The more general geometries of GTR, such as the projective, conformal, affine and metric structures, are characterized as geometric fields over a manifold M. Typically, such geometric fields are mathematically represented as cross sections of a fiber bundle. A geometric object at a point $p \in M$, where M denotes the spacetime manifold, is typically a 'Taylor series' approximation of some finite degree k at $p \in M$ of a map either into or out of M. The dimension n of M is usually 4. In general, the geometric objects at p of a given type form a differentiable manifold $F(M_p)$ of some finite dimension ℓ called the *fiber* over $p \in M$. Each of these manifolds is diffeomorphic to a manifold $\mathbb{F}$ called the *typical fiber*. The disjoint union of the fibers $F(M_p)$ form a differential manifold $F(M)$ of dimension $n + \ell$. The structure

$$\mathcal{F}(M) = \langle F(M), \pi, M, \mathbb{F} \rangle ,$$

is called a fiber bundle over the base space M, where $\pi : F(M) \to M$ is the projection map which maps every geometric object in $F(M_p)$ into $p \in M$. A cross section $\mathcal{F}(M)$ is a map $\sigma : M \to F(M)$ such that $\pi \circ \sigma = \mathrm{id}_M$. A geometric-object field[11] of type $\mathbb{F}$ is a cross section of $\mathcal{F}(M)$.

To repeat, in the case of GTR, physical or theoretic coordinate systems cannot be introduced in the initial stages of the modeling process even with the aid of ad hoc assumptions, because the circumstances that permitted such assumptions in the pre-GTR case do not obtain. At the initial stage of the modeling process, therefore, only the manifold M has been introduced. Nothing that represents a physical entity has been postulated. There are neither physical fields nor material bodies. It follows that the charts $\{(U_\alpha, x_\alpha)\}$ cannot be either theoretic or physical charts; consequently, they must be purely formal. Because the manifold M does not represent anything that is physical, in particular, its points do not represent the events of physical spacetime, and because the only requirement on its differentiable structure is that it be smoother than the differentiable structure of physical spacetime, one is free to assume that the differentiable structure of M is C^∞.

The same geometric field structure, represented by means of a cross section, can be placed over M in many ways. One precisely describes this situation by means of a formal-active transformation. A formal-active transformation is determined by a diffeomorphism $f : M \to M$. Consider an arbitrary point $p \in M$ and suppose

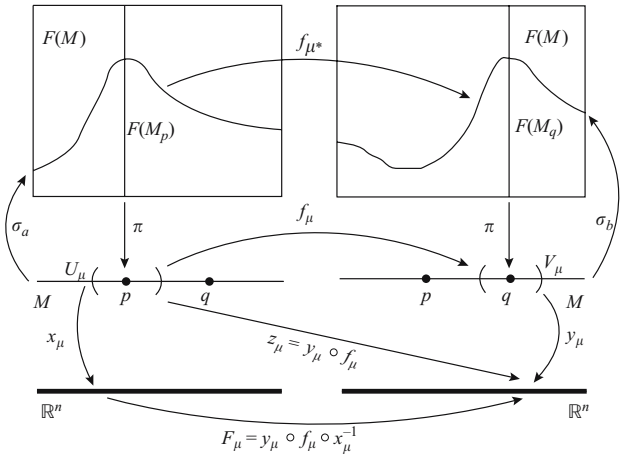

FIGURE 9.5. Active to passive.

$f(p) = q \in M$. Then, geometric objects at $p \in M$ are mapped into corresponding geometric objects at $q \in M$ by the 'drag-along' process. For example, a tangent vector $V_p \in T(M_p)$ is 'pushed forward' to a vector $f_* V_p \in T(M_{f(p)})$. This operation can be extended to geometric-object fields.

Under a formal-active transformation, the points of M remain fixed but the geometric field is actually moved or 'dragged along' to another location of the manifold. In the context of a model for spacetime, the fields that represent the geometric structure of the physical world are represented as cross sections of various bundles over M. Two models of the geometry that are determined by different cross sections σ_a and σ_b are regarded as equivalent provided that there is a diffeomorphism of the manifold M that carries σ_a into σ_b; that is, just in case the models are related by a formal-active transformation (model diffeomorphism). The placing of the same geometric structure in two different locations with respect to the base manifold M by means of a formal-active transformation, is illustrated in Figure 9.5. By proposition 2.5 a formal-active transformation is equivalent to a formal-passive transformation. According to the equivalent, albeit conceptually quite different *passive* point of view, nothing is actually being moved; instead a mere formal re-labeling of coordinates of the base manifold M has taken place. As a consequence, a formal-active transformation, that is, a model diffeomorphism, has no physical consequences whatsoever; the physics of the situation remains entirely unchanged.

2.5 Earman and Norton on the hole argument

Someone who believes in the doctrine of spacetime or manifold substantivalism asserts, broadly speaking, that the spacetime manifold, together with its points, is a physically real entity which is endowed with physically real differential-topological relations between its points. Sentences like 'Denote by M the 4-dimensional differentiable manifold of spacetime' occur frequently in the literature and they seem

to support a manifold substantivalist conception for they suggest that there exists a physical entity that is an aspect of spacetime, a sort of container, with specific physical properties. Of course, as was discussed earlier, the notion of a spacetime manifold endowed with a differentiable structure is, in the first instance of the modeling process, a conceptual and/or semantic necessity; for without such a concept, one would be unable to introduce the bundle of non-degenerate, symmetric, second order tensors and one could therefore not introduce the metric tensor as a cross section of this bundle. Moreover, one would also be unable to discuss the world paths of material bodies and their equations of motion. The question to be raised here is whether manifold substantivalism leads to 'radical local indeterminism', as was suggested by Earman and Norton, and to what extent it is or is not necessary to accept the substantivalist position. Earman (1986, 1989), Norton (1987, 1988, 1989, 1992) and Earman and Norton (1987) have argued in the context of their discussion of Einstein's hole argument[12] that a spacetime-manifold substantivalist is faced with *radical local indeterminism*. In his book *World enough and Spacetime*, Earman (1989, 190) suggests that on an active construal of the hole argument "there are in the offing many different metrics that, according to manifold substantivalism, predict objectively different properties of space-time points." For example, Earman suggests that under one diffeomorphism of the metric "points p and q are relatively lightlike," while under another diffeomorphism "they are relatively spacelike, which leads to the conflicting prediction that p and q can and cannot be connected by a nonbroken light ray."

The first difficulty with Earman and Norton's account is a purely formal one. The kind of diffeomorphism involved in their active construal of the hole construction is a purely *formal diffeomorphism* which occurs in discussions about diffeomorphically equivalent models of spacetime theories and statements about the uniqueness of the solution of the Initial Value problem for GTR up to a diffeomorphism. Such a formal-active transformation is described with respect to a purely formal system of coordinates. Earman and Norton evidently think that the passive viewpoint is *substantively* different from the active viewpoint in this context, in the sense that an active, as opposed to a passive construal, will lead to conflicting predictions. But as was shown earlier, a formal-active transformation, that is, a model diffeomorphism, is equivalent to a formal-passive transformation. Therefore, the situation that Earman describes, cannot even arise on purely formal grounds: an equivalent formal-passive transformation results in a mere formal re-labeling of the coordinate descriptions of the manifold points and therefore cannot have any effect on lightlike, spacelike or timelike relations.

The second difficulty concerns their conception of manifold substantivalism according to which the points of M represent the actual or possible events of physical spacetime and constitute the *relata* of lightlike, spacelike or timelike relations which also reside in M. However, the events of physical spacetime and the relations between them do not lie in the base manifold M; they lie elsewhere. They must be regarded as aspects of the cross section in the bundle space $F(M)$ that represent the geometry of physical spacetime. Spacetime points representing events of physical spacetime and the timelike, lightlike and spacelike relations between them, exist at the *field-body*

level of reality which is mathematically represented in terms of cross sections in the bundle space $F(M)$. We shall call such points *field-body points* in order to distinguish them from the *base-manifold points* of M.

An example might clarify the situation. Suppose a physical radar coordinate chart based on some observer A has been set up. Then a point P in its domain can be physically designated by referring to its coordinates $x^i_A(P)$. Suppose that the spacetime metric $g_{Aij}(x^i_A)dx^i_A \otimes dx^j_A$ has been measured in at least a portion of the chart's domain. Suppose further that P_2 is in the future of P_1. Then one can *physically* designate a double cone open set of points, namely, all those points that are both in the future of P_1 and in the past of P_2. It is clear that the collection of all possible physically designated open double cone sets is a topology that is isomorphic to the manifold topology. The *physically* designated spacetime points are singled out by their relations to physical fields and bodies. Such spacetime points exist, therefore, at the field-body level of reality and are the field-body points. It should also be noted that the field-body points have topological relations in addition to those they share with the base-manifold points of M. In particular, field-body points are also timelike, lightlike or spacelike related.

With regard to the issue of spacetime substantivalism, one may adopt one of two points of view:

Field-Body Relationalism: Only fields, material bodies and their world paths, and relations between these entities are physically real. On the other hand, the spacetime manifold, its points and the differential-topological relations between them do not exist physically but only conceptually; they merely provide a semantic framework that is necessary for the theoretical activity of modeling the world.

Spacetime Substantivalism: In addition to the field-body level of reality, there exists a 'container', the spacetime manifold, and this manifold, its points and the manifold differential-topological relations are physically real.

Recall Weyl's remark, that the metric becomes "freed from space; it becomes an existing *field* within the remaining structure-less space. Through this, space as form of appearance contrasts more clearly with its real content: the content is measured after the form is arbitrarily related to coordinates." Weyl's description is compatible with both field-body relationalism and spacetime substantivalism.

It describes field-body relationalism if the remaining structure-less space or the space as form of appearance does not denote a physically real container but merely denotes, as it were, a conceptual scaffolding that is used by the theorist to model the contents of the world. The contents of the world reside at the field-body relationalist level, and include such things as the dynamics and interactions of the various physical entities of the world including the physical procedures, such as physical coordinates used to survey the world.

It describes spacetime substantivalism if the remaining structure-less space does denote a physically real container. However, as *form* of appearances it cannot then itself be among the appearances. That is, the substantivalist must accept the fact that, although this manifold, its points and the manifold differential-topological relations

are physically real, the spacetime manifold plays no role other than to constrain the topology at the field-body relationalist level to be compatible with the manifold topology. The manifold points have no causal efficacy whatsoever since the relations of timelikeness, spacelikeness and lightlikeness occur only at the field-body relationalist level. We have no direct epistemic access to the manifold points and their topological relations. We have knowledge of the manifold differential-topological structure only because the structure is duplicated at the field-body relationalist level. Furthermore, any formal executed model diffeomorphism merely re-positions the physical fields and bodies in the container without changing any of the field-body relations and hence without changing any physical causal relations.

Whether or not one adopts the field-body relationalist or the spacetime substantivalist position, there is no incompatibility between the fact that the solutions to Einstein's field equations are unique only up to an active diffeomorphism and the statement that the spacetime metric can be uniquely determined empirically. Geometric-object fields are locally described by a system of components (descriptors) with respect to a local coordinate chart. In the context of pure mathematics the local coordinates are merely stipulated or assumed and the corresponding components are called *formal* descriptors. On the other hand, the measurement of a physical geometric-object field is the empirical determination of its *physical* descriptors, that is, its system of components with respect to a local *physical* radar-station coordinate chart. *Physical descriptors* are *covariant* with respect to a change from one physical coordinate system to another physical coordinate system. The measurement situation can, of course, be described with respect to various local, *formal* coordinate charts. It should be clear that under a formal-passive transformation, the *physical* descriptors of the geometric-object field being measured, are *invariants*. Since a formal-passive transformation is equivalent to a formal-active transformation the *physical* descriptors are also invariant from an active point of view. Since a formal-active transformation merely re-positions everything, including the local physical coordinate systems and the geometric-object fields, none of the field-body relations (such as *physical* descriptors) are changed by a formal-active transformation that relates two formal solutions in the context of Einstein's hole argument. Any formal-active transformation f that transforms a geometric field Ξ into another geometric field Ξ^f will also transform the chosen physical coordinate system Σ into another coordinate system Σ^f such that Ξ^f will have the same field-body relation with respect to Σ^f as Ξ has to Σ; that is, the physical descriptors of the geometric field with respect to some physical coordinate system are *invariants* under formal-active transformations.

These and earlier considerations make it clear that a spacetime-manifold substantivalist is *not* faced with 'radical local indeterminism'. The points of the container, though physically real according to the spacetime substantivalist, have no causal efficacy and we have no direct epistemic access to them. The relations 'timelike', 'spacelike' or 'lightlike' occur only at the field-body relationalist level and *not* at the container level. Earman's (1986, 182) suggestion that "…[p]hysics has accomplished what Leibniz' Principle of Sufficient Reason could not" does not seem correct. It would seem that in arguing against the substantivalist view we can essentially do

no better than Leibniz. We can point out that the manifold differential-topological relations are duplicated at the field-body relationalist level and that since one has direct epistemic access only to the field-body relations, the postulation of the physical existence of the manifold is ontologically gratuitous.

It is interesting to note in this context that the use of $\mathfrak{G}$-structures for the characterization of geometric fields such as the metric field, may lead to important insights with respect to these fields. For example, a $\mathfrak{G}$-structure may be flat or non-flat; but it can never vanish. Consequently, fields characterizable as $\mathfrak{G}$-structures do not vanish. The reason is that the n^k-co-frames are geometric objects that cannot vanish at any point $p \in M$ because the first order part of the approximation must be invertible since the n^k-co-frame is the Taylor approximation of a local diffeomorphism $h : U \to \mathbb{R}^n$ which satisfies $h(p) = \vec{0}$; consequently, *no unoccupied* spacetime points can exist. This, it would seem, undercuts a version of Leibnizian relationalism which, as (Friedman, 1983, 217) describes, "places constraints on the *ontology* of our space-time theories" and which "wishes to limit the domain over which the quantifiers of our theories range to the set of physical events, that is, the set of space-time points that are actually occupied".

In light of this one might adopt the following attitude: The fact that the metric field has been, to use Weyl's terminology, freed from the manifold, and the fact that the metric field cannot be removed (that is, does not vanish anywhere), and finally the fact that the manifold, postulated by the substantivalist, plays no essential physical role other than to constrain the topology at the field-body relationalist level, suggests that we reject manifold substantivalism and adopt instead a *pure* field-body relationalist position which regards the manifold as a mere mental construct or conceptual scaffolding. That is, the *true* locations and differential-topological relations (as well as other relations) between them are inherent features of the geometric structure itself and not that of the base manifold M. This leads to the question of how to correctly characterize these true relations as opposed to a purely formal, and somewhat misleading, characterization in terms of the purely formal differential structure of the base manifold M. We will not answer this question here but merely point out that it can be shown that certain geometric structures, such as, for example, the affine structure or symmetric linear connection, are so rich that they 'carry' their own differentiable-manifold structure; that is, these geometries determine a complete atlas of coordinate charts with respect to which the geometry itself can be described in a self-referential manner.

3 FIELD-BODY RELATIONALISM: THE NECESSITY OF GEOMETRIC FIELDS

Those who argue for the conventional character of the law of inertia from ontological considerations concerning the nature of spacetime structure and/or for their relationalist character from a Leibnizian-Machian view of motion — in which relative motion must be understood as relative motion of bodies with respect to each other — advance the theses that what counts as a standard of no-acceleration or free

motion is not dictated by a physically real and causally efficacious inertial structure of spacetime.

Weyl was a vigorous opponent of such an ontological view and argued that inertial effects are a consequence of the inertial structural field which he called the *guiding field* and that pure body relationalism is incoherent within the context of GTR.[13] To emphasize the necessity for a physically real and causally efficacious inertial structure of spacetime, a structure Weyl called the *guiding field*, Weyl (1924, 1949) devised the following paradox:

> Incidentally, without a world structure the concept of relative motion of several bodies has, as the postulate of general relativity shows, no more foundation than the concept of absolute motion of a single body. Let us imagine the four-dimensional world as a mass of plasticine traversed by individual fibers, the world lines of material particles. Except for the condition that no two world lines intersect, their pattern may be arbitrarily given. The plasticine can then be continuously deformed so that not only one but all fibers become vertical straight lines. Thus no solution of the problem is possible as long as in adherence to the tendencies of Huyghen and Mach one disregards the structure of the world. But once the inertial structure of the world is accepted as the cause for the dynamical inequivalence of motions, we recognize clearly why the situation appeared so unsatisfactory …. Hence the solution is attained as soon as we dare to *acknowledge the inertial structure as a real thing that not only exerts effects upon matter but in turn suffers such effects*.[14]

Let us analyze this example using the concept of the microsymmetry group[15] of a geometric structure at an event $p \in M$. A *microsymmetry* of an acceleration field or of a directing field (or other structural fields) at an event $p \in M$ is a local diffeomorphism of a neighbourhood of $p \in M$ which leaves p fixed and preserves the field at the event $p \in M$. The set of microsymmetries at $p \in M$ form the microsymmetry group at $p \in M$. Consider a spacetime manifold equipped only with a differentiable structure, the plasticine of Weyl's example. Our spacetime does not have a connection defined on it. In such a world it is possible to define curves and paths and their elements. However, there are no preferred curves or paths. Since there is only the differentiable structure, one may apply any diffeomorphism; consequently, the microsymmetry group at any event p is an infinite-parameter group isomorphic to the group of all invertible formal power series in four variables. If there is no post-differentiable topological geometric field in the neighbourhood of the event $p \in M$, then all of these infinite parameters may be chosen freely within rather broad limits. Clearly then, given an infinite number of parameters, one can straighten out an arbitrary pattern of world lines (fibers) in the neighbourhood of any event. Now suppose that there exists a post-differentiable topological geometric field, namely, the projective structure, or geodesic directing field. Then the microsymmetry group that preserves that structure is a 20-parameter Lie group. Thus instead of an infinity of degrees of freedom, only twenty degrees of freedom may be used to actively deform the

neighbouring region of spacetime. The fact that only a finite number of parameters are available prevents an arbitrary realignment of the worldlines of material bodies in the neighbourhood of any given event.

Other post-differential topological geometric field are similarly restrictive. For example, the microsymmetry group of the conformal structure which determines the causal structure of spacetime permits 7 degrees of freedom (6 Lorentz transformations and a dilatation) and permits four more degrees of freedom in second order.

Weyl's plasticine example shows that the Leibnizian view of relative motion, namely the view according to which all motion must be defined as motion relative to bodies, is self-defeating in GTR. The fact that a stationary, homogeneous elastic sphere will, when set in rotation, bulge at the equator and flatten at the poles is well known. According to Weyl, this phenomenon is to be accounted for in the following way. The complete physical system consisting of both the body and the local inertial-gravitational field is not the same in the two situations. The cause of the effect is the state of motion of the body with respect to the local gravitational field and is not, indeed as Weyl's plasticine example shows, cannot be the state of motion of the body relative to other bodies. To attribute the effect, as Einstein and Mach did, to the rotation of the body with respect to other *bodies* in the universe, is to endorse a remnant of the unjustified monopoly of the older body ontology, namely, the sovereign right of material bodies to play the role of physically real and acceptable causal agents. Our ontology must be extended to include, according to Weyl, causally efficacious *geometrical structural fields* (*geometrische Strukturfelder*) and this leads to the following modern reformulation of Newton's laws of motion.

First consider the inhomogeneous character of the transformation laws for 4-acceleration and 3-acceleration. Furthermore, for the sake of simplicity consider the case of linear acceleration only. Denote by M the n-dimensional, C^∞ spacetime manifold. A *curve* in M is a map $\gamma : \mathbb{R} \to M$ and a *path* in M is an equivalence class $\xi = [\gamma]$ of such maps any two of which are related by an invertible parameter transformation $\mu : \mathbb{R} \to \mathbb{R}$. For convenience, in the discussion of curve and path elements[16] at some particular point $p \in M$, attention is restricted to those curves which satisfy $\gamma(0) = p$ and to those parameter transformations which satisfy $\mu(0) = 0$.

A *curve element* of order k at $p \in M$ is an equivalence class $j_0^k\gamma$ of curves through p which have the same Taylor expansion with respect to some (and hence every) coordinate chart $(U, x)_p$ up to and including order k at $0 \in \mathbb{R}$. A *path element* of order k at $p \in M$ is an equivalence class of paths $j_p^k\xi$ consisting of all paths corresponding to curves in $j_0^k\gamma$, where $\gamma \in \xi$.

A second-order curve element $j_0^2\gamma$ has local coordinates γ_1^i and γ_2^i called 4-velocity and 4-acceleration, respectively, and given by

$$\gamma_1^i = \frac{d}{d\lambda} x^i \circ \gamma(0),$$

$$\gamma_2^i = \frac{d^2}{d\lambda^2} x^i \circ \gamma(0).$$

A second-order path element $j^2_p\xi$ has local coordinates ξ^α_1 and ξ^α_2 called 3-velocity and 3-acceleration, respectively, and given by

$$\xi^\alpha_1 = \left.\frac{dx^\alpha \circ \gamma}{dx^0 \circ \gamma}\right|_p,$$
$$\xi^\alpha_2 = \left.\frac{d^2x^\alpha \circ \gamma}{(dx^0 \circ \gamma)^2}\right|_p.$$

The transformation laws of the coordinates γ^i_1 (4-velocity) and γ^i_2 (4-acceleration) under a change of coordinate chart from $(U,x)_p$ to $(\bar{U},\bar{x})_p$, follow from the pointwise definition of γ^i_1 and γ^i_2. From

$$\bar{\gamma}^i(\lambda) = \bar{x}^i(\gamma(\lambda)) = \bar{X}^i(x(\gamma(\lambda))) = \bar{X}^i(\gamma^i(\lambda)),$$

where $\bar{X}^i = \bar{x}^i \circ x^{-1}$, one obtains

$$\bar{\gamma}^i_1 = \bar{X}^i_j \gamma^j_1$$

and

$$\bar{\gamma}^i_2 = \bar{X}^i_j \gamma^j_2 + \bar{X}^i_{jk} \gamma^j_1 \gamma^k_1, \tag{3}$$

where $\bar{X}^i_j$ and $\bar{X}^i_{jk}$ are the first and second partial derivatives of $\bar{X}^i(x^i)$ at $x^i(p)$.

Under a change of coordinate chart from $(U,x)_p$ to $(\bar{U},\bar{x})_p$, the transformation laws for the coordinates ξ^α_1 (3-velocity) and ξ^α_2 (3-acceleration), are given by

$$\bar{\xi}^\alpha_1 = \frac{\bar{X}^\alpha_0 + \bar{X}^\alpha_\beta \xi^\beta_1}{\bar{X}^0_0 + \bar{X}^0_\gamma \xi^\gamma_1},$$

and

$$\begin{aligned}\bar{\xi}^\alpha_2 = {} & \frac{\bar{X}^\alpha_\beta \xi^\beta_2 + \bar{X}^\alpha_{\rho\sigma} \xi^\rho_1 \xi^\sigma_1 + 2\bar{X}^\alpha_{0\rho} \xi^\rho_1 + \bar{X}^\alpha_{00}}{(\bar{X}^0_0 + \bar{X}^0_\gamma \xi^\gamma_1)^2} \\ & - \frac{\bar{X}^0_\beta \xi^\beta_2 + \bar{X}^0_{\rho\sigma} \xi^\rho_1 \xi^\sigma_1 + 2\bar{X}^0_{0\rho} \xi^\rho_1 + \bar{X}^0_{00}}{(\bar{X}^0_0 + \bar{X}^0_\gamma \xi^\gamma_1)^2} \bar{\xi}^\alpha_1.\end{aligned} \tag{4}$$

The transformation laws (3) and (4) for the 4-acceleration and 3-acceleration are linear in the acceleration variables but they are *not* homogeneous. The inhomogeneity of the transformation laws entails that an acceleration that is zero with respect

to one coordinate system is not zero with respect to another coordinate system. It follows that given *only* a differential-topological structure, there is no standard for zero acceleration. Since the terms that are independent of the acceleration depend on both the spacetime location and on the corresponding velocity, it is necessary to specify a standard for zero acceleration that depends on those variables, namely, position and velocity. That is, it is necessary to specify fields, either a second order geodesic acceleration field $\Gamma^i_2(x^i, \gamma^i_1)$ or a second order geodesic directing field $\Pi^\alpha_2(x^i, \xi^\alpha_1)$.

The transformation law for an acceleration field can be obtained from (3) by replacing $\bar{\gamma}^i_2$ by $\bar{A}^i_2(\bar{x}^i, \bar{\gamma}^i_1)$ and γ^j_2 by $A^j_2(x^i, \gamma^i_1)$ to yield

$$\bar{A}^i_2(\bar{x}^i, \bar{\gamma}^i_1) = \bar{X}^i_j A^j_2(x^i, \gamma^i_1) + \bar{X}^i_{jk}\gamma^j_1\gamma^k_1.$$

The important special case for which $A^i_2(x^i, \gamma^i_1)$ is a geodesic acceleration field corresponds to a $\mathfrak{G}$-structure, namely, the affine structure. For this special case the function $A^i_2(x^i, \gamma^i_1)$ is denoted by $\Gamma^i_2(x^i, \gamma^i_1)$ and is given by

$$\Gamma^i_2(x^i, \gamma^i_1) = -\Gamma^i_{jk}(x^i, \gamma^i_1)\gamma^j_1\gamma^k_1.$$

The transformation law for the affine structure is given by

$$\bar{\Gamma}^i_2(\bar{x}^1, \bar{\gamma}^i_1) = \bar{X}^i_j \Gamma^j_2(x^i, \gamma^i_1) + \bar{X}^i_{jk}\gamma^j_1\gamma^k_1.$$

The transformation law for a directing field $\Xi^\alpha_2(x^i, \xi^\alpha_1)$ can be obtained from (4) by replacing $\bar{\xi}^\alpha_2$ by $\bar{\Xi}^\alpha_2(\bar{x}^i, \bar{\xi}^\alpha_1)$ and ξ^β_2 by $\Xi^\beta_2(x^i, \xi^\alpha_1)$. The important special case for which $\Xi^\alpha_2(x^i, \xi^\alpha_1)$ is *cubic* in the (3)-velocity variables ξ^α_1 corresponds to a $\mathfrak{G}$-structure, namely, the projective structure. For this special case, the function $\Xi^\alpha_2(x^i, \xi^\alpha_1)$ is denoted by $\Pi^\alpha_2(x^i, \xi^\alpha_1)$ and is given by

$$\begin{aligned}\Pi^\alpha_2(x^i, \xi^\alpha_1) = {} & \xi^\alpha_1[\Pi^0_{\rho\sigma}(x^i, \xi^\alpha_1)\xi^\rho_1\xi^\sigma_1 + 2\Pi^0_{0\rho}(x^i, \xi^\alpha_1)\xi^\rho_1 + \Pi^0_{00}(x^i, \xi^\alpha_1)] \\ & - [\Pi^\alpha_{\rho\sigma}(x^i, \xi^\alpha_1)\xi^\rho_1\xi^\sigma_1 + 2\Pi^\alpha_{0\rho}(x^i, \xi^\alpha_1)\xi^\rho_1 + \Pi^\alpha_{00}(x^i, \xi^\alpha_1)],\end{aligned} \tag{5}$$

where the projective coefficients satisfy $\Pi^i_{jk} = \Pi^i_{kj}$ and $\Pi^j_{jk} = 0$. The transformation law for the projective structure is given by

$$\begin{aligned}\bar{\Pi}^\alpha_2(\bar{x}^i, \bar{\xi}^\alpha_1) = {} & \frac{\bar{X}^\alpha_\beta \Pi^\beta_2(x^i, \xi^\alpha_1) + \bar{X}^\alpha_{\rho\sigma}\xi^\rho_1\xi^\sigma_1 + 2\bar{X}^\alpha_{0\rho}\xi^\rho_1 + \bar{X}^\alpha_{00}}{(\bar{X}^0_0 + \bar{X}^0_\gamma\xi^\gamma_1)^2} \\ & - \frac{\bar{X}^0_\beta \Pi^\beta_2(x^i, \xi^\alpha_1) + \bar{X}^0_{\rho\sigma}\xi^\rho_1\xi^\sigma_1 + 2\bar{X}^0_{0\rho}\xi^\rho_1 + \bar{X}^0_{00}}{(\bar{X}^0_0 + \bar{X}^0_\gamma\xi^\gamma_1)^2}\bar{\xi}^\alpha_1.\end{aligned} \tag{6}$$

The differences

$$\gamma_2^i - \Gamma_2^i(x^i, \gamma_1^i) \tag{7}$$

and

$$\xi_2^\alpha - \Pi_2^\alpha(x^i, \xi_1^\alpha) \tag{8}$$

then transform linearly and homogeneously; consequently, the vanishing or non-vanishing of these relative accelerations is coordinate independent.[17]

According to the modern re-formulation of the laws of motion, the law of inertia asserts the existence of a unique projective structure:

The Law of Inertia: There exists on spacetime a unique projective structure Π_2 or equivalently, a unique geodesic directing field Π_2.

Thus formulated, the law of inertia is an empirical law. It is falsifiable, for if there exist at least two sets of particles each of which is governed by a distinct *geodesic* directing field Π_2 and Π_2', then particles belonging to the two distinct directing-field sets may be identified and in turn may be used to measure in any chosen local neighbourhood of spacetime the two distinct projective structures. This discovery procedure is a non-circular, coordinate and frame independent epistemically effective procedure which makes use of a purely local differential topological criterion for geodesicity.[18] *Free motion* is defined with reference to the projective structure Π_2 as follows:

Definition of Free Motion: A possible or actual material body is in a state of free motion during any part of its history just in case the corresponding segment of its world path is a solution path of the differential equation determined by the unique projective structure of spacetime

The law of inertia and the definition of free motion together constitute a modern reformulation of Newton's first law of motion. Newton's second law of motion may be reformulated as follows:

The Law of Motion: With respect to any coordinate system, the world line path of a possible or actual material body satisfies an equation of the form

$$m(\xi_2^\alpha - \Pi_2^\alpha(x^i, \xi_1^\alpha)) = F^\alpha(x^i, \xi_1^\alpha),$$

where m is a scalar constant characteristic of the material body called its inertial mass and $F^\alpha(x^i, \xi_1^\alpha)$ is the 3-force acting on the body.

The components ξ_2^α of the 3-acceleration can be thought of as the kinematic descriptors of a material body. On the other hand, the components of the geodesic directing field $\Pi_2^\alpha(x^i, \xi_1^\alpha)$ are *field* quantities. The difference $(\xi_2^\alpha - \Pi_2^\alpha(x^i, \xi_1^\alpha))$ denotes the

components of a coordinate independent *field-body relation*. The transformation law for 3-forces is linear and *homogeneous* since the inhomogeneous terms cancel for the transformation law of the difference $(\xi_2^\alpha - \Pi_2^\alpha(x^i, \xi_1^\alpha))$.[19]

Note that the law of motion makes explicit use of the unique projective structure Π_2 on spacetime. The law of motion, therefore, depends ontologically on the law of inertia; consequently, it is impossible to derive the law of inertia from the law of motion. Specifically, it is not the case that the first law is derivable from the second law as a special instance, a claim that is made even in excellent textbooks.

It is also the case that 3-forces are directly measurable since the directing fields and the projective structure are directly measurable; consequently, 3-forces are *physically* real.

It is interesting to note in this context a theorem proved by Coleman and Korté (1989) which says: If a second order directing field $\Xi_2(x^i, \xi_1^\alpha)$ is a polynomial with respect to its 3-velocity variables ξ_1^α in every coordinate chart, then it is necessarily geodesic.

This result establishes that forces are necessarily non-polynomial with respect to the 3-velocity variables. Recently Coleman and Korté (1999) have shown how this result makes explicit and clarifies an essential difference between pre-relativistic and relativistic theories: In contrast with the relativistic case for which the projective structure is unique, in pre-relativistic physics, *all* of the physical directing fields then known were, in present terminology, geodesic; that is, they all corresponded to projective structures and hence were geometrizable. It is essentially because of this circumstance that it was not possible in pre-relativistic theories to demarcate by means of a local criterion the boundary between forces and geometry in a non-stipulative manner, and hence to formulate Newton's laws of motion in a non-circular way.

NOTES

[1] (Einstein, 1954, 289) remarks:

> Moreover I believed that I could show on general considerations that a law of gravitation invariant with respect to arbitrary transformations of coordinates was inconsistent with the principle of causality. These were errors of thought which cost me two years of excessively hard work, until I finally recognized them as such at the end of 1915, and after having ruefully returned to the Riemannian curvature, succeeded in linking the theory with the facts of astronomical experience.

See also Stachel (1986, 1987) and Lanczos (1972) for further historical comments and related issues.

[2] In the final analysis of his hole problem, Einstein tried to cope with the need to secure physical determinism by claiming that all physical observations reduce to the observation of coincidences and that such coincidences are invariant under the (formal) diffeomorphisms under consideration.

[3] See for example, (Bergmann and Komar, 1980, 230), (Wald, 1984, 438), Adler et al. (1975, 279), Hawking and Ellis (1974, 230), D'Inverno (1998, 178).

[4] If f is a map, then the maps $f_{\vdash}$ and $f^{\dashv}$ respectively denote the image map and inverse image map determined by f. A good account of this somewhat nonstandard notation may be found in Porteous (1969).

[5] This container need not be accorded the status of a physically real entity, rather it may be regarded as a mere formal scaffolding that is required for the modeling process. See the discussion of field-body relationalism presented below.

[6] See Arnowitt et al. (1962).

[7] An account of this *local* formulation of the equivalence may also be found in the text by Wald (1984, 439).

[8] Of course, most of the known exact solutions of the field equations exhibit considerable symmetry precisely because such solutions are easier to find.

[9] In the case of pre-GTR, the Galilean theories, the use of purely formal coordinates is not necessary because of the assumption pertaining to the flatness of the geometric structures. This assumption then permits the use of theoretic and physical coordinates that are adapted to the flat geometric structures. Nevertheless, even in the context of pre-GTR, a proper theoretical account of the epistemology of geometry still requires an analysis of the physical measurement process including an analysis of physical coordinates with respect to these theoretic coordinates. (See, Coleman and Korté (1995a,b))

[10] The relation between the physical coordinates and the geometric structure in GTR is described by a complicated system of functions $g_{ij}(x^i)$. The corresponding relationship in classical mechanics is described by a small number of constants; for example, $g_{\alpha\beta}(x^\alpha) = \delta_{\alpha\beta}$ for the spatial metric.

[11] The usage of the terms 'geometric-object field' and 'geometric field' is such that the first includes the second. By a 'geometric field', we mean a $\mathfrak{G}$-structure, such as a Riemannian, conformal, affine or projective structure. Besides these structures, the term 'geometric-object field' also includes such fields as equation-of-motion structures for massive monopoles and the electromagnetic field.

[12] For other discussions on the topic of the hole argument see Butterfield (1989), Maudlin (1988, 1990), Nerlich (1991, 1994), Stachel (1989).

[13] For a discussion of Weyl's philosophy of spacetime see Korté (1981) and Coleman and Korté (2001).

[14] Weyl's emphasis.

[15] For a treatment of the concept of *microsymmetry* see Coleman and Korté (1984); for an analysis of G-*structures* see Coleman and Korté (1981), Coleman and Korté (1993).

[16] For a treatment of *curve elements* and *path elements* and equation-of-motion structures in terms of the *jet-formalism* and for a new formulation of the *laws of motion*, see Coleman and Korté (1980, 1981, 1982, 1984).

[17] In (7) and (8) we have assumed that the field that determines the zero acceleration is geodesic; however, the universality of free fall motion could in principle be determined by a non-geodesic acceleration or directing field. (See Coleman and Korté (1984).)

[18] See Coleman and Korté (1980, 1982, 1984, 1987, 1989, 1990).

[19] Sklar (1974, 229–233) has suggested that the problem of absolute acceleration arises because we tend to think of acceleration as a *dyadic* relation: something accelerates either with respect to some observable or unobservable entity. Sklar suggests that there is an alternative way to think of absolute acceleration, which, if adopted by the relationalist, will avoid the traditional relationalist difficulties concerning absolute acceleration. He proposes that we think of acceleration as a *monadic* relation so that "the expression '*A* is absolutely accelerated' is a complete assertion, as is, for example, '*A* is red'. . .." It should be clear from the foregoing, however, that Sklar's suggestion is incoherent within the context of GTR. The inhomogeneity of the transformation law of the 3-acceleration entails that an acceleration that is zero with respect to one coordinate system is not zero with respect to another coordinate system. Consequently, absolute nonacceleration conceived of as a monadic property is not a well defined concept because it is not a coordinate independent notion. A monadic property which can be transformed away by means of a passive coordinate transformation can hardly represent a brute, inexplicable fact about the world. However, the difference $\xi_2^\alpha - \Pi_2^\alpha(x^i, \xi_1^\alpha)$ transforms linearly and homogeneously; consequently, the vanishing or non-vanishing of these *field-body relations is* coordinate independent.

In theories prior to the advent of GTR, the affine and projective structures were flat. Moreover, the habit of using coordinates adapted to these structures made it difficult to appreciate the important role

these structures have, since the components that describe them, vanish in an adapted coordinate system; consequently these structures were inconspicuous. In such a context, one could perhaps be seduced into thinking that acceleration could be understood as a monadic relation.

REFERENCES

Ronald Adler, Maurice Bazin, and Menahem Schiffer. *Introduction to General Relativity*. McGraw-Hill Book Company, New York, 2 edition, 1975.

R. Arnowitt, S. Deser, and C. W. Misner. The dynamics of general relativity. In L. Witten, editor, *Gravitation: An Introduction to Current Research*. John Wiley, New York, 1962.

Peter G. Bergmann and Arthur Komar. The phase space formulation of general relativity and approaches toward its canonical quantization. In A. Held, editor, *General Relativity and Gravitation*, volume 1, pages 227–254. Plenum Press, New York, 1980.

J. Butterfield. The hole truth. *British Journal for the Philosophy of Science*, 40:1–28, 1989.

R. A. Coleman and H. Korté. Jet bundles and path structures. *The Journal of Mathematical Physics*, 21(6):1340–1351, 1980.

R. A. Coleman and H. Korté. Spacetime G-structures and their prolongations. *The Journal of Mathematical Physics*, 22(11):2598–2611, 1981.

R. A. Coleman and H. Korté. The status and meaning of the laws of inertia. In *The Proceedings of the Biennial Meeting of the Philosophy of Science Association*, pages 257–274, Philadelphia, 1982.

R. A. Coleman and H. Korté. Constraints on the nature of inertial motion arising from the universality of free fall and the conformal causal structure of spacetime. *The Journal of Mathematical Physics*, 25(12): 3513–3526, 1984.

R. A. Coleman and H. Korté. Any physical, monopole, equation-of-motion structure uniquely determines a projective inertial structure and an $(n-1)$-force. *The Journal of Mathematical Physics*, 28(7):1492–1498, 1987.

R. A. Coleman and H. Korté. All directing fields that are polynomial in the $(n-1)$-velocity are geodesic. *The Journal of Mathematical Physics*, 30(5):1030–1033, 1989.

R. A. Coleman and H. Korté. Harmonic analysis of directing fields. *The Journal of Mathematical Physics*, 31(1):127–130, 1990.

R. A. Coleman and H. Korté. An empirical, purely spatial criterion for the planes of F-simultaneity. *Foundations of Physics*, 24(4):417–437, 1991.

R. A. Coleman and H. Korté. On attempts to rescue the conventionality thesis of distant simultaneity in STR. *Foundations of Physics Letters*, 5(6):535–571, 1992a.

R. A. Coleman and H. Korté. The relation between the measurement and Cauchy problems of GTR. In Humitaka Sato and Takashi Nakamura, editors, *The Sixth Marcel Grossmann Meeting on General Relativity*, pages 97–119. World Scientific, 1992b. Printed version of an invited talk presented at the meeting held in Kyoto, Japan, 23–29 June 1991.

R. A. Coleman and H. Korté. Why fundamental structures are of first or second order. *Journal of Mathematical Physics*, 35(4):1803–1818, 1993.

R. A. Coleman and H. Korté. A new semantics for the epistemology of geometry I, Modeling spacetime structure. *Erkenntnis*, 42:141–160, 1995a.

R. A. Coleman and H. Korté. A new semantics for the epistemology of geometry II, Epistemological completeness of Newton-Galilei and Einstein-Maxwell theory. *Erkenntnis*, 42:161–189, 1995b.

R. A. Coleman and H. Korté. Geometry and forces in relativistic and pre-relativistic theories. *Foundations of Physics Letters*, 12(2):147–163, 1999.

R. A. Coleman and H. Korté. Hermann Weyl: Mathematician, Physicist, Philosopher. In Erhard Scholz, editor, *Hermann Weyl's Raum – Zeit – Materie and a General Introduction to His Scientific Work*, volume 30 of *Deutsche Mathematiker-Vereinigung Seminar*, pages 161–386. Birkhäuser, Basel, 2001.

Ray D'Inverno. *Introducing Einstein's Relativity*. Clarendon Press, Oxford, 1998.

J. Earman. *A Primer on Determinism*. Reidel, Dordrecht, 1986.
J. Earman. *World Enough and Space-Time*. MIT, Cambridge Massachusetts, 1989.
J. Earman and J. Norton. What price spacetime substantivalism? *British Journal for the Philosophy of Science*, 38:515–525, 1987.
J. Ehlers, R. A. E. Pirani, and A. Schild. The geometry of free fall and light propagation. In L. O' Raifeartaigh, editor, *General Relativity, Papers in Honour of J. L. Synge*, pages 64–84. Clarendon Press, Oxford, 1972.
A. Einstein. Notes on the origin of the general theory of relativity. In *Ideas and Opinions*, pages 285–290. Bonanza Books, New York, 1954.
M. Friedman. *Foundations of Space-Time Theories*. Princeton University Press, Princeton, 1983.
S. W. Hawking and G. F. R. Ellis. *The Large Scale Structure of Space-Time*. Cambridge University Press, Cambridge, 1974.
H. Korté. *A Realist Interpretation of the Causal-Inertial Structure of Spacetime*. PhD thesis, University of Western Ontario, London, Ontario, August 1981.
C. Lanczos. Einstein's path from special to general relativity. In L. O' Raifeartaigh, editor, *General Relativity, Papers in Honour of J. L. Synge*, pages 5–19. Clarendon Press, Oxford, 1972.
T. Maudlin. The essence of space-time. In *PSA 1988*, volume 2 of *PSA Proceedings*, pages 82–91. Philosophy of Science Association, 1988.
T. Maudlin. Substances and space-time: What Aristotle would have said to Einstein. *Stud. Hist. Phil. Science*, 21(4):531–561, 1990.
G. Nerlich. How euclidean geometry has misled metaphysics. *The Journal of Philosophy*, 88(4):169–189, April 1991.
G. Nerlich. *What spacetime explains*. Cambridge University Press, 1994.
J. Norton. Einstein, the hole argument and the reality of space. In J. Forge, editor, *Measurement, Realism and Objectivity*, pages 153–188. Reidel, 1987.
J. Norton. The hole argument. In *PSA 1988*, volume 2 of *PSA Proceedings*, pages 56–64. Philosophy of Science Association, 1988.
J. Norton. Coordinates and covariance: Einstein's view of space-time and the modern view. *Foundations of Physics*, 19(10):1215–1263, 1989.
J. Norton. The physical content of general covariance. In J. Eisenstaedt and A. J. Kox, editors, *Studies in the History of General Relativity*, volume 3 of *Einstein Studies*, pages 281–315. Birkhäuser, Boston, 1992. Based on the Proceedings of the 2nd International Conference on the History of General Relativity, Luminy, France, 1988.
I. R. Porteous. *Topological Geometry*. Van Nostrand Reinhold Company, London, 1969. Reprinted in 1972.
L. Sklar. *Space, Time, and Spacetime*. University of California Press, Berkeley, 1974.
J. Stachel. The meaning of general covariance: The hole story. John Earman, Allen I. Janis, Gerald J. Massey, and Nicholas Rescher, editors, *Philosophical Problems of the Internal and External Worlds: Essays on the Philosophy of Adolf Grünbaum*, pages 129–160, University of Pittsburgh Press, Pittsburgh-Konstanz series in the philosophy and history of science, Pittsburgh, 1993.
J. Stachel. What a physicist can learn from the discovery of general relativity. In R. Ruffini, editor, *Proceedings of the fourth Marcel Grossmann Meeting on General Relativity*, pages 1857–1862, Amsterdam, 1986. Elsevier.
J. Stachel. How Einstein discovered general relativity: a historical tale with some contemporary morals. In M. A. H. MacCallum, editor, *General Relativity and Gravitation, Proceedings of the 11th International Conference on General Relativity and Gravitation*, pages 200–208, Cambridge, 1987. Cambridge University Press.
J. Stachel. Einstein's search for general covariance, 1912–1915. In D. Howard and J. Stachel, editors, *Einstein and the History of General Relativity*, volume 1 of *Einstein Studies*, pages 63–100. Birkhäuser, Boston, 1989. Based on the Proceedings of the 1986 Osgood Hill Conference, North Andover, Massachusetts 8–11 May 1986.
R. M. Wald. *General Relativity*. Chicago Press, Chicago, 1984.

H. Weyl. Massenträgheit und Kosmos. Ein Dialog. *Die Naturwissenschaften*, 12:197–204, 1924. Reprinted in Weyl (1968).

H. Weyl. On the foundations of infinitesimal geometry. *Bulletin of the American Mathematical Society*, 35:716–725, 1929. Reprinted in Weyl (1968).

H. Weyl. *Philosophy of Mathematics and Natural Science*. Princeton University Press, Princeton, 1949.

H. Weyl. *Gesammelte Abhandlungen*, volume I–IV. Springer Verlag, Berlin, 1968. Edited by K. Chandrasekharan.

H. Weyl. *Riemanns geometrische Ideen, ihre Auswirkung und ihre Verknüpfung mit der Gruppentheorie*. Springer-Verlag, 1988. Posthumous publication; edited by K. Chandrasekharan.

ITAMAR PITOWSKY*

10. QUANTUM MECHANICS AS A THEORY OF PROBABILITY

ABSTRACT

We develop and defend the thesis that the Hilbert space formalism of quantum mechanics is a new theory of probability. The theory, like its classical counterpart, consists of an algebra of events, and the probability measures defined on it. The construction proceeds in the following steps: (a) Axioms for the algebra of events are introduced following Birkhoff and von Neumann. All axioms, except the one that expresses the uncertainty principle, are shared with the classical event space. The only models for the set of axioms are lattices of subspaces of inner product spaces over a field K. (b) Another axiom due to Solèr forces K to be the field of real, or complex numbers, or the quaternions. We suggest a probabilistic reading of Solèr's axiom. (c) Gleason's theorem fully characterizes the probability measures on the algebra of events, so that Born's rule is derived. (d) Gleason's theorem is equivalent to the existence of a certain finite set of rays, with a particular orthogonality graph (Wondergraph). Consequently, all aspects of quantum probability can be derived from rational probability assignments to finite "quantum gambles". (e) All experimental aspects of entanglement- the violation of Bell's inequality in particular- are explained as natural outcomes of the probabilistic structure. (f) We hypothesize that even in the absence of decoherence, macroscopic entanglement can very rarely be observed, and provide a precise conjecture to that effect. We discuss the relation of the present approach to quantum logic, realism and truth, and the measurement problem.

1 INTRODUCTION

Discussions of the foundations of quantum mechanics have been largely concerned with three related foundational questions which are often intermingled, but which I believe should be kept apart:

1. A semi-empirical question: Is quantum mechanics complete? In other words, do we have to supplement or restrict the formalism by some additional assumptions?
2. A mathematical-logical question: What are the constraints imposed by quantum mechanics on its possible alternatives? This is where all the famous "no-hidden-variables" theorems belong.
3. A philosophical question: Assuming that quantum mechanics is complete, what then does it say about reality?

* Department of Philosophy, The Hebrew University, e-mail: Itamarp@vms.huji.ac.il

W. Demopoulos and I. Pitowsky (eds.), Physical Theory and its Interpretation, 213–240.

By *quantum mechanics* I mean the Hilbert space formalism, including the dynamical rule for the quantum state given by Schrödinger's equation, Born's rule for calculating probabilities, and the association of measurements with Hermitian operators. These elements seem to me to be the core of the (nonrelativistic) theory.

I shall be concerned mainly with the philosophical question. Consequently, for the purpose of this paper the validity and completeness of the Hilbert space formalism is assumed. By making this assumption I do not wish to prejudge the answer to the first question. It seems to me dogmatic to accept the completeness claim, since no one can predict what future theories will look like. At the same time I think it is also dogmatic to reject completeness. Present day alternatives to quantum mechanics, be they collapse theories like GRW [1], or non-collapse theories like Bohm's [2], all suffer from very serious shortcomings.

However, one cannot ignore the strong philosophical motivation behind the search for alternatives. These are, in particular, two conceptual assumptions, or perhaps dogmas that propel this search: The first is J. S. Bell's dictum that the concept of measurement should not be taken as fundamental, but should rather be defined in terms of more basic processes [3]. The second assumption is that the quantum state is a real physical entity, and that denying its reality turns quantum theory into a mere instrument for predictions. This last assumption runs very quickly into the measurement problem. Hence, one is forced either to adopt an essentially non-relativistic alternative to quantum mechanics (e.g. Bohm without collapse, GRW with it); or to adopt the baroque many worlds interpretation which has no collapse and assumes that all measurement outcomes are realized.

In addition, the first assumption delegates secondary importance to measurements, with the result that the uncertainty relations are all but forgotten. They are accepted as empirical facts, of course; but after everything is said and done we still do not know why it is impossible to measure position and momentum at the same time. In Bohm's theory, for example, the commutation relations are adopted by fiat even on the level of individual processes, but are denied any fundamental role in the theory.

My approach is traditional and goes back to Heisenberg, Bohr and von Neumann. It takes the uncertainty relations as the centerpiece that demarcates between the classical and quantum domain. This position is mathematically expressed by taking the Hilbert space, or more precisely, the lattice of its closed subspaces, as the structure that represents the "elements of reality" in quantum theory. The quantum state is a derived entity, it is a device for the bookkeeping of probabilities. The general outlook presented here is thus related to the school of quantum information theory, and can be seen as an attempt to tie it to the broader questions of interpretation. I strive to explain in what way quantum information is different from classical information, and, perhaps why.

The main point is that the Hilbert space formalism is a "logic of partial belief" in the sense of Frank Ramsey [4]. In such a logic one usually distinguishes between possible "states of the world" (in Savage's terminology [5]), and the probability function. The former represent an objective reality and the latter our uncertainty about it. In the quantum context possible states of the world are represented by the closed subspaces of the Hilbert space while the probability is derived from the $|\psi\rangle$ function

by Born's rule. In order to avoid confusion between the objective sense of *possible state* (subspace), and $|\psi\rangle$- which is also traditionally called the state- we shall refer to the subspaces as *events*, or *possible events*, or *possible outcomes* (of experiments). To repeat, my purpose is to defend the position that the Hilbert space formalism is essentially a new theory of probability, and to try to grasp the implications of this structure for reality.

The initial plausibility of this approach stems from the observation that quantum mechanics uses a method for calculating probabilities which is different from that of classical probability theory[1]. Moreover, in order to force quantum probability to conform to the classical mold we have to add objects (variables, events) and dynamical laws over and above those of quantum theory. This state of affairs calls for a philosophical analysis because the theory of probability is a theory of inference and, as such, is a guide to the formation of rational expectations.

The relation between the above stated purpose and the completeness assumption should be stressed again. We can always avoid the radical view of probability by adopting a non-local, contextual hidden variables theory such as Bohm's. But then I believe, the philosophical point is missed. It is like taking Steven Weinberg's position on space-time in general relativity: There is no non-flat Riemannian geometry, only a gravitational field defined on a flat space-time that appears as if it gives rise to geometry [9–11]. I think that Weinberg's point and also Bohm's theory are justified only to the extent that they yield new discoveries in physics (as Weinberg certainly hoped). So far they haven't.

Jeffrey Bub was my thesis supervisor over a quarter of a century ago, and from him I have Iearnt the mysteries of quantum mechanics and quantum logic [12]. For quite a while our attempts to grasp the meaning of the theory diverged, but now seem to converge again [13]. It is a great pleasure for me to contribute to this volume in honor of a teacher and a dear friend.

2 THE EVENT STRUCTURE

2.1 Impossibility, certainty, identity, and the non contextuality of probability

Traditionally a theory of probability distinguishes between the set of possible events (called the algebra of events, or the set of states of Nature, or the set of possible outcomes) and the probability measure defined on them. In the Bayesian approach what constitutes a possible event is dictated by Nature, and the probability of the event represents the degree of belief we attach to its occurrence. This distinction, however, is not sharp; what is possible is also a matter of judgment in the sense that an event is judged impossible if it gets probability zero in all circumstances. In the present case we deal with physical events, and what is impossible is therefore dictated by the best available physical theory. Hence, probability considerations enter into the structure of the set of possible events. We represent by 0 the equivalence class of all events which our physical theory declares to be utterly impossible (never occur, and

therefore always get probability zero) and by 1 what is certain (always occur, and therefore get probability one).

Similarly, the *identity* of events which is encoded by the structure also involves judgments of probability in the sense that *identical events always have the same probability.* This is the meaning of accepting a structure as an algebra of events *in a probability space*. An important example is the following: Consider two measurements A, B, which can be performed together, so that $[A, B] = 0$; and suppose that A has the possible outcomes $a_1, a_2, \ldots, a_k$, and B the possible outcomes $b_1, b_2, \ldots, b_r$. Denote by $\{A = a_i\}$ the event "the outcome of the measurement of A is a_i", and similarly for $\{B = b_j\}$. Now consider the identity:

$$\{B = b_j\} = \bigcup_{i=1}^{k} (\{B = b_j\} \cap \{A = a_i\}) \tag{1}$$

This is the distributivity rule which holds in this case as it also holds in all classical cases. This means, for instance, that if A represents the roll of a die with six possible outcomes and B the flip of a coin with two possible outcomes, then Eq (1) is trivial. Consequently the probability of the left hand side of Eq (1) equals the probability of the right hand side, for every probability measure.

In the quantum mechanical context this observation has further implications. If A, B, C, are observables such that $[A, B] = 0$, and $[B, C] = 0$ but $[A, C] \neq 0$. Then the identity

$$\bigcup_{i=1}^{k} (\{B = b_j\} \cap \{A = a_i\}) = \{B = b_j\} = \bigcup_{i=1}^{l} (\{B = b_j\} \cap \{C = c_i\}) \tag{2}$$

holds, where $c_1, c_2, \ldots, c_l$ are the possible outcomes of C. By the rule *Identical events always have the same probability* we conclude that the probabilities of all three expressions in Eq (2) are equal. This is *the principle of the non-contextuality of probability*. There is a large body of literature which attempts to justify this principle[2]. For why should we apply the same probability to $\{B = b_j\}$ in the A, B context as in the B, C context? If this is a good question in the quantum domain it should be an equally good question in the classical regime. For consider Eq (1) with A representing the throw of a die, and B the flip of a coin. Now think of two contexts: In one we just flip the coin without rolling the die; in the other we do both. Why should the probability of $\{B = b_j\}$ be the same in both contexts? (regardless of our judgment about the dependence, or independence of the events). By the very act of putting the outcomes of the two procedures "coin flipping" and "die throwing" in the same *probability* space (the product space) we are ipso facto assuming Eq (1) as an identity in a probability space which implies equality of probabilities. Although routinely made, this assumption ultimately represents an empirical judgment. Counterexamples are hard to come by, and are usually quite contrived.

My proposal to take the Hilbert space formalism as a Ramsey type logic of partial belief involves the same commitment. Hence, in the following I *assume* that the 0

of the algebra of subspaces represents impossibility (zero probability in all circumstances) 1 represents certainty (probability one in all circumstances), and the identities such as Eq (1) and Eq (2) represent identity of probability in all circumstances. This is the sense in which the lattice of closed subspaces of the Hilbert space is taken as an algebra of events. I take these judgments to be natural extensions of the classical case; a posteriori, they are all justified empirically.

2.2 The axioms

In their 1936 seminal paper "The logic of quantum mechanics" Garrett Birkhoff and John von Neumann [19] formulated the quantum logical program. Their strategy was to take the following steps:

1. Identify the quantum structure which is the analogue of the event structure of classical statistical mechanics.
2. Distill a set of principles underlying this structure and formulate them as axioms.
3. Show that the quantum structure is, in some sense, THE model of the axioms.

Birkhoff and von Neumann identified the quantum event structure (which they called "quantum logic") as the algebra of closed subspaces of a Hilbert space. In the rest of this section I shall review the efforts to accomplish steps 2 and 3 of their program, that is, begin with the axioms and generate the structure. The elements in the structure we shall refer to as "events" , or "outcomes" (meaning outcomes of gambles or of measurements) or sometimes loosely as "propositions" (meaning propositions that describe the events). Notice that the axioms below are shared by both classical and quantum systems, with the exception of the last axiom. It should also be noted that I do not claim that this structure is *logic* in the same sense that the predicate calculus or intuitionistic logic are. (Nor do I think that Birkhoff and von Neumann made such a claim).[3] A proposition that describes a possible event in a probability space is of a rather special kind. It is constrained by the requirement that there should be a viable procedure to determine whether the event occurs, so that a gamble that involves it can be unambiguously decided. This means that we exclude many propositions. For example, propositions that describe past events of which we have only a partial record, or no record at all. We also exclude undecidable mathematical propositions such as the continuum hypothesis, and many other propositions that form a part of the standard conception of logic. Our structure is "logic" only insofar as it is the event component of a "logic of partial belief".

We use small Latin letters $x, y, \ldots$, to designate events, and denote by L the totality of events. $\cap$ stands for intersection, $\cup$ for union, and implication is denoted by $\leq$. Finally, $x^{\perp}$ denotes the complement of x. The certain event is denoted by 1 and the null event by 0.

These are the axioms:

S1 $x \leq x$.

S2 If $x \leq y$ and $y \leq z$ then $x \leq z$.

S3 If $x \leq y$ and $y \leq x$ then $x = y$.

S4 $0 \leq x \leq 1$

S5 $x \cap y \leq x$, and $x \cap y \leq y$, and if $z \leq x$ and $z \leq y$ then $z \leq x \cap y$.

S6 $x \leq x \cup y$, and $y \leq x \cup y$, and if $x \leq z$ and $y \leq z$ then $x \cup y \leq z$.
O1 $(x^{\perp})^{\perp} = x$
O2 $x \cap x^{\perp} = 0$ and $x \cup x^{\perp} = 1$
O3 $x \leq y$ implies $y^{\perp} \leq x^{\perp}$.
O4 Orthomodularity if $x \leq y$ then $y = x \cup (y \cap x^{\perp})$.
Axiom O4 is sometimes replaced by a stronger axiom:
O4* Modularity if $x \leq z$ then $x \cup (y \cap z) = (x \cup y) \cap z$.

The axioms S1-S6, O1-O4 are true in the classical system of propositional logic or, more precisely, in the Lindenbaum-Tarski algebra of such a logic, when we interpret the operations as logical connectives. The rest of the axioms are more specific to the physical context.

H1 Atomism: If $x \lneqq y$ then there is an atom p such that $p \leq y$ and $p \not\leqslant x$. Here by an *atom* we mean an element $0 \neq p \in L$ such that $x \leq p$ entails $x = 0$ or $x = p$.

H2 Covering property: For all atoms p and all elements x if $x \cap p = 0$ then $x \leq y \leq x \cup p$ entails $y = x$ or $y = x \cup p$.

Atomism and the covering property are introduced to ensure that every element of the lattice is a union of atoms. The atoms, whose existence is guaranteed by H1, are maximally informative propositions. In the classical case they correspond to points in the phase space (or rather, singleton subsets of phase space); in the quantum case they correspond to one dimensional subspaces of the Hilbert space.[4]

H3 Completeness: if $S \subset L$ then $\cup_{a \in S} a$ and $\cap_{a \in S} a$ exist.

Usually we do not assume such a strong axiom in the classical physical case. There, the algebra of possible events is the σ-algebra of Lebesgue measurable subsets of phase space, which is assumed to be closed only under countable unions and intersections. However, axiom H3 is *consistent* with the classical physical event space. It is known that in some models of set theory every set of reals is Lebesgue measurable [24]. In such models H3 will automatically be satisfied for the Lebesgue algebra in phase space. This means that no substantial difference between the classical and quantum case arises from H3.

The one single axiom that separates the quantum from the classical domain is

H4 Irreducibility: If z satisfies for all $x \in L$ $x = (x \cap z) \cup (x \cap z^{\perp})$ then $z = 0$ or $z = 1$.

This last axiom is non-classical in the following sense: there is only one Boolean algebra which is irreducible, the trivial algebra $\{0, 1\}$. In classical physics the set of events is a large Boolean algebra. In fact, it is *totally reducible*: for all x and all z we have $x = (x \cap z) \cup (x \cap z^{\perp})$.

So consider the case

$$x \neq (x \cap z) \cup (x \cap z^{\perp}) \tag{3}$$

The intuitive meaning of Eq (3) is that the events x and z are *incompatible*, that is, cannot be the outcomes of a single experiment. Thus, axiom H4 is the formal expression of indeterminacy. Later we shall see how Eq (3) entails a more familiar uncertainty relation between the probabilities of x and z. For the sake of illustration,

at this stage, consider the case in which x and z are atoms. One implication of Eq (3) is that *there are non orthogonal atoms*. So consider some measurement in which x is the *actual* outcome, and the other possible outcomes are x', x'',…etc., all orthogonal to x, so that z is not among them. This means that after the measurement is performed we gain no knowledge as to whether z is the case or not. This state of affairs would not be very surprising were it not for the fact that x and z are atomic events; but in this case it seems to imply that *there is no fact of the matter as to whether z is the case or not*. In other words, no certain record about the possible outcome z is obtainable, in principle, while we perform the x measurement. By "fact" I mean here, and throughout, a *recorded* fact, an actual outcome of a measurement. Restricting the notion of "fact" in this way should not be understood, at this stage, as a metaphysical thesis about reality. It is simply the concept of "fact" that is analytically related to our notion of "event", in the sense that only a recordable event can potentially be the object of a gamble. Later, in section 4.1 and in the last section we shall come back to this issue, when we discuss the implications of the theory to the structure of reality.

2.3 Representations and the gap

In the classical case we assume that *for all x and z the following holds* $x = (x \cap z) \cup (x \cap z^{\perp})$. This makes the lattice L an atomic Boolean algebra. More specifically $(L, 0, 1, \leq, \cap, \cup, \perp)$ is isomorphic to the Boolean algebra of the subsets of the set of all atoms, with the usual Boolean operators, with 1 the set of all atoms and 0 the null set.

The representation theorem for quantum systems is more complicated, in this case $(L, 0, 1, \leq, \cap, \cup, \perp)$ is isomorphic to the lattice of subspaces of a vector space with a scalar product, more specifically:

1. There is a division ring K (field whose product is not necessarily commutative), with involutional automorphism $* : K \to K$, that is, for all $\alpha, \beta \in K$ $\alpha^{**} = \alpha$, $(\alpha + \beta)^* = \alpha^* + \beta^*$, $(\alpha\beta)^* = \beta^*\alpha^*$.
2. There's a (left) vector space V over K.
3. There's a Hermitian form $<,>: V \times V \to K$ satisfying for all $u, v, w \in V$, and $\alpha, \beta \in K$
   $< \alpha u + \beta v, w >= \alpha < u, v > + \beta < v, w >$,
   $< u, \alpha v + \beta w >=< u, v > \alpha^* + < u, w > \beta^*$,
   $< u, v >=< v, u >^*$,
   $< u, u >= 0$ if and only if $u = 0$.

Let $X \subset V$ be a subspace, let $X^{\perp} = \{v \in V; < u, v >= 0 \; \forall u \in X\}$ then $X^{\perp}$ is also a subspace. If $X = X^{\perp\perp}$ we shall say that X is *closed*, then we have $V = X \oplus X^{\perp}$. The representation theorem asserts that L is isomorphic to the lattice of closed subspaces of V, in other words $L \simeq \{X \subset V; X = X^{\perp\perp}\}$. The operation $\cap$ is just subspace intersection, and $X \cup Y = (X^{\perp} \cap Y^{\perp})^{\perp}$.

The proof of this representation theorem has essentially two parts. The first is the classical representation theorem for projective geometries which goes back to the middle of the 19th century.[5] An irreducible, atomic, complete, lattice with a complementation $\perp$ that satisfies O2, and which is modular (O4*) is a projective geometry.

The traditional result on the coordinatization of projective geometries yields the field K and the vector space V over it. Adding the stronger conditions on $\perp$, in particular O4, enabled Birkhoff and von Neumann to derive the inner product structure on V. Note that so far we have not introduced any explicit physical assumption, or even a probabilistic assumption, save perhaps the indeterminacy implicit in H4. Nevertheless, we see that the principle of superposition (that is, the fact that V is a linear space) already presents itself.

In both the classical and the quantum cases some additional assumptions are needed to obtain the actual models. In the quantum case the construction will be completed if we are able to infer that $K = \mathbb{C}$ (the field of complex numbers) on the basis of a probabilistic or a physically intuitive axiom. At least we would like to force K to be either the field of real numbers, or the complex numbers, or the quaternions. In these cases the inner product of a non-zero vector by itself $< u, u >$ is a positive real number, and Gleason's theorem describes the probabilistic structure. This is a gap in the argument which has been closed to a certain extent (in the case of infinite dimensional Hilbert spaces) by the work of Solèr [27, 28]. It is hoped that a reasonable more straightforward probabilistic or information theoretic axiom (such as a constraint on tensor-like products) will close the gap even more tightly.

2.4 Solèr's axiom and theorem

The best result known in this direction involves a geometric axiom. It is the celebrated theorem of Maria Pia Solèr which applies in case the lattice is infinite dimensional. The extra axiom connects a projective geometric concept (harmonic conjugation) to the orthogonality structure. Recall that a projective geometry is associated with the lattice in the following way: Every atom is a *point* every pair of atoms generates *a projective line* and every triple of atoms which are not colinear determine a *projective plane*. Let x and y be two atoms then the line through them is $x \cup y$. Suppose that z is another atom on this line, $z \leq x \cup y$, then we construct a fourth point $w \leq x \cup y$ on the line which is called the harmonic conjugate of z relative to x and y -denoted by $w = \mathcal{H}(z; x, y)$- as follows (Figure 10.1): Let $u \not\leq x \cup y$ be arbitrary and let $v \leq x \cup u$, $v \neq x, u$. Denote by

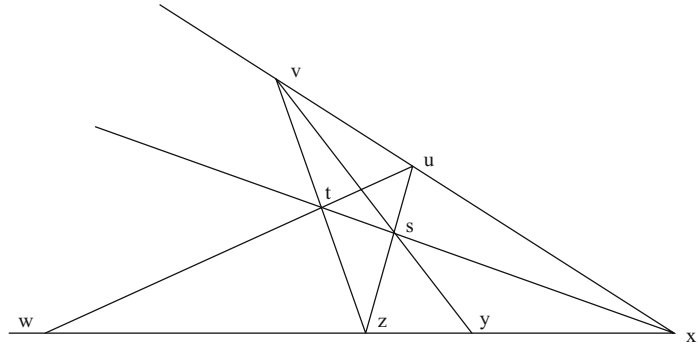

FIGURE 10.1. Harmonic conjugation.

$s = (z \cup u) \cap (y \cup v)$ and $t = (x \cup s) \cap (z \cup v)$ then $w \doteqdot \mathcal{H}(z; x, y) \doteqdot (u \cup t) \cap (x \cup y)$. The harmonic conjugate is unique (that is, independent of the choice of u and v). It is a basic construction in projective geometry, closely related to the definitions of the algebraic operations in the field K (realized as the projective line).

Soler's axiom may be phrased as follows:

SO If x and y are orthogonal atoms then there is $z \leq x \cup y$ such that $w = \mathcal{H}(z; x, y)$ is orthogonal to z. In other words, $\mathcal{H}(z; x, y) = z^{\perp} \cap (x \cup y)$.

Intuitively, such a z bisects the angle between x and y, that is, defines $\sqrt{2}$ in the field K. Soler's proved

Theorem 1 *If L is infinite dimensional and satisfies* SO *then K is $\mathbb{R}$ or $\mathbb{C}$ or the quaternions.*

In fact she proved a stronger result, assuming only that there is an infinite sequence of orthogonal atoms $\{x_i\}_{i \in \mathbb{N}}$ such that $x = x_i$ and $y = x_{i+1}$ satisfy SO for every $i = 1, 2, \ldots$ The axiom SO may be given a probabilistic interpretation in the spirit Ramsey as we shall see subsequently.

3 PROBABILITY MEASURES: GLEASON'S THEOREM, WONDERGRAPH AND SOLÈR'S AXIOM.

3.1 Gleason's theorem

Assume that the set of possible events (or possible measurement outcomes, or propositions) is the lattice $L = L(\mathbb{H})$ of subspaces of a *real or complex* Hilbert space $\mathbb{H}$. For simplicity, we shall concentrate on the finite dimensional case. Our aim is to tie this structure to probabilities, and by doing so to provide further evidence that the elements of L can be seen as representing quantum events. Moreover, we shall see how the traditional features and "paradoxes" of quantum mechanics are expressed and resolved in the quantum probabilistic language.

First a few words to connect measurements and outcomes in the more traditional view with the present notations. Here we shall be concerned with measurements that have a finite set of possible outcomes. Let A be an observable (a Hermitian operator) with n distinct possible numerical real values (the eigenvalues of A) $\alpha_1, \alpha_2, \ldots, \alpha_n$. With each value corresponds an event $x_i = \{A = \alpha_i\}$ meaning "the outcome of a measurement of A is α_i" We identify this event with the subspace of $\mathbb{H}$ spanned by the eigenvectors of A having the eigenvalue α_i. The events x_i are pair-wise orthogonal elements of L. The sub lattice that $x_1, x_2, \ldots, x_n$ generate is a finite Boolean algebra which we shall denote by $\mathcal{B} = \langle x_1, x_2, \ldots, x_n \rangle$. In case n is the dimension of the space $\mathbb{H}$ each one of the events x_i is an atom and the observable A is said to be maximal.

Subsequently we shall identify any observable A with the Boolean algebra $\langle x_1, x_2, \ldots, x_n \rangle$ generated by its outcomes. Note that this is an unusual identification. It means that we equate the observables A and $f(A)$, whenever f is a one-one function defined on the eigenvalues of A. This step is justified since we are interested

in *outcomes* and not their labels, and hence in such a "scale free" concept of observable. (It is like replacing the numbers $1, 2, \ldots, 6$ on the face of a die by the numbers $2, 3, \ldots, 7$ respectively.) The converse is also true, with each orthogonal set of elements $x_1, x_2, \ldots, x_n$ of L there corresponds an observable whose eigenspaces include these elements.

Probability measures which are definable on L were characterized many years ago in case $n = \dim \mathbb{H} \geq 3$. Since every set of n orthogonal atoms represents the outcomes of a possible measurement, and since they are all the possible outcomes we are motivated to introduce

Definition 1 *Suppose that $\mathbb{H}$ is of a finite dimension n over the complex or real field. A real function P defined on the atoms in L is called a state (or alternatively, a probability function) on $\mathbb{H}$ if the following conditions hold*

1. $P(0) = 0$, *and* $P(y) \geq 0$ *for every element* $y \in L$.
2. *If* $x_1, x_2, \ldots, x_n$ *is an orthogonal set of atoms then* $\sum_{j=1}^{n} P(x_j) = 1$.

The probability of every lattice element $y \in L$ is then fixed since it is a union of a set of orthogonal atoms $y = x_1 \cup \ldots \cup x_r$, so that $P(y) = \sum_{j=1}^{r} P(x_j)$. A complete description of the possible states is given by Gleason's theorem [29]:

Theorem 2 *Given a state P on a space of dimension* ≥ 3 *there is an Hermitian, non negative operator W on H, whose trace is unity, such that* $P(x) = < \vec{x}, W\vec{x} >$ *for all atoms* $x \in L$, *where* $<,>$ *is the inner product, and* $\vec{x}$ *is a unit vector along x. In particular, if some* $x_0 \in L$ *satisfies* $P(x_0) = 1$ *then* $P(x) = \left| < \vec{x}_0, \vec{x} > \right|^2$ *for all* $x \in L$ *(Born's rule).*

With the obvious conditions on convergence the above definition and theorem generalize to the infinite dimensional case. The remarkable feature exposed by Gleason's theorem is that the event structure dictates the quantum mechanical probability rule. It is one of the strongest pieces of evidence in support of the claim that the Hilbert space formalism is just a new kind of probability theory. The quantum structure is in this sense much more constrained than the classical formalism. The structure of the phase space of a classical system does not gratly restrict the type of probability measures that can be defined on it. The probability measures which are actually used in classical statistical mechanics are introduced mostly by fiat or, in any case, are very hard to justify.

Gödel [30] said in a different context : "A probable decision about the truth [of a new axiom] is possible ... inductively by studying its "success" . Success here means fruitfulness in consequences in particular "verifiable" consequences, i.e., consequences demonstrable without the axiom". Importing this insight from the mathematical domain to the present physical domain we can see how the set of axioms for the structure, most of which are shared with classical probability, give rise to the quantum mechanical probabilistic structure which is otherwise left a mystery.

3.2 Finite gambles and uncertainty

So far we have dealt with the lattice L in its entirety, and with everywhere defined probability functions. The standard conceptions of Bayesian probability theory make do, initially at least, of finite probability spaces. A canonical situation handled by this theory is that of a gamble. In the words of Ramsey: "The old-established way of measuring a person's belief" by proposing a bet, and seeing what are the lowest odds which he will accept, is "fundamentally sound" [4]. Our gambles will likewise be finite and consist of four steps

1. A *single* physical system is prepared by a method known to everybody.
2. A *finite* set $\mathcal{M}$ of incompatible measurements, each with a finite number of possible outcomes, is announced by the bookie. The agent is asked to place bets on the possible outcomes of each one of them.
3. One of the measurements in the set $\mathcal{M}$ is chosen by the bookie and the money placed on all other measurements is promptly returned to the agent.
4. The chosen measurement is performed and the agent gains or looses in accordance with his bet on that measurement.

There are two reasons to concentrate on finite gambles of this kind. First, to avoid over idealization; for it is hard to imagine someone betting on the outcomes of all possible measurements (perhaps writing an IOU for each one of them). Secondly, and more importantly, the infinite idealization blurs the important fact that indeterminacy, and all other "strange" results associated with quantum theory, are fundamentally combinatorial. The non-classical behavior of the probabilities is already forced by a finite number of events and the relations among them.

Recall that each measurement is identified with the Boolean algebra generated by its possible outcomes in L: $\mathcal{B} = \langle x_1, x_2, \ldots, x_m \rangle$ (the x_i's may not be atomic in case $\mathcal{B}$ is not a maximal measurement). So a gamble $\mathcal{M}$ is just a set of such algebras $\mathcal{M} = \{\mathcal{B}_1, \mathcal{B}_2, \ldots, \mathcal{B}_k\}$. We do not assume that the gambler knows quantum theory. All she is aware of is the logical structure which consists of these sets of outcomes. In particular, *she recognizes identities*, and the cases where the same outcome is shared by more than one experiment. By acting according to the standards of rationality the gambler will assign probabilities to the outcomes. To see this, assume that $P(x \mid \mathcal{B})$ is the probability assigned by the agent to the outcome x in measurement $\mathcal{B}$, where $\mathcal{B} \in \mathcal{M}$ and $x \in \mathcal{B}$.

RULE 1: *For each measurement $\mathcal{B} \in \mathcal{M}$ the function $P(\cdot \mid \mathcal{B})$ is a probability distribution on $\mathcal{B}$.*

This follows directly from classical probability theory. Recall that after the third stage in the quantum gamble the agent faces a bet on the outcome of a single measurement. The situation at this stage is essentially the same as a tossing of a coin or a casting of a die. Hence, the probability values assigned to the possible outcomes of the chosen measurement should be coherent.

RULE 2: *If $\mathcal{B}_1, \mathcal{B}_2 \in \mathcal{M}$,* and $y \in \mathcal{B}_1 \cap \mathcal{B}_2$ *then* $P(y \mid \mathcal{B}_1) = P(y \mid \mathcal{B}_2)$.

The rule asserts the non contextuality of probability, discussed in section 2.1. Suppose that $\mathcal{B}_1 = \langle x_1, x_2, \ldots, x_m \rangle$ and $\mathcal{B}_2 = \langle z_1, z_2, \ldots, z_r \rangle$ then $y \in \mathcal{B}_1 \cap \mathcal{B}_2$

implies that $(x_1 \cap y) \cup \ldots \cup (x_m \cap y) = y = (z_1 \cap y) \cup \ldots \cup (z_r \cap y)$. Rule 2, therefore, follows from this identity between events, and the principle that identical events in a probability space have equal probabilities.

To take the discussion closer to the lattice theoretic conception consider finite subsets of events, $\Gamma \subset L$.

Definition 2 *Two propositions x and y of* Γ *are compatible if* $x = (x \cap y) \cup (x \cap y^{\perp})$ *and* $y = (y \cap x) \cup (y \cap x^{\perp})$. *A state (or probability function)* Γ *is a real function P on* Γ *such that*

a. $P(x) \geq 0$ *for all* $x \in \Gamma$
b. $P(x^{\perp}) = 1 - P(x)$ *whenever* $x, x^{\perp} \in \Gamma$
c. $P(x \cup y) + P(x \cap y) = P(x) + P(y)$ *whenever x and y are compatible and* $x, y \in \Gamma$.

Such probability functions defined over finite subsets of events in the lattice are the subject of our study. Note that we do not put any requirements on such P's apart from the three conditions a, b, c, in the definition. In particular, probability functions on Γ are not constrained to be induced by quantum mechanical states. The relation between this definition and the gambles introduced previously is clear. Given any gamble $\mathcal{M}$ as above the set of events is $\Gamma = \mathcal{B}_1 \cup \mathcal{B}_2 \cup \ldots \cup \mathcal{B}_k \subset L$. Every probability function which follows *RULE 1* and *RULE 2* satisfy the conditions a, b, c, in definition 2.

As a simple example which demonstrates an uncertainty relation consider the following quantum gamble $\mathcal{M}$ consisting of seven incompatible measurements (Boolean algebras), each generated by its three possible atomic outcomes:

$$\langle x_1, x_2, y_2 \rangle\,,\ \langle x_1, x_3, y_3 \rangle\,,\ \langle x_2, x_4, x_6 \rangle\,,\ \langle x_3, x_5, x_7 \rangle\,,$$
$$\langle x_6, x_7, y \rangle\,,\ \langle x_4, x_8, y_4 \rangle\,,\ \langle x_5, x_8, y_5 \rangle$$

Note that some of the outcomes are shared by two measurements, these are denoted by the letter x. The other outcomes each belong to a single algebra, and are denoted by a y. The orthogonality relations among the generators are depicted in the *orthogonality graph* in Figure 10.2, which is a part of Kochen and Specker's famous "cat's cradle"[31]. Each node in the graph represents an outcome, two nodes are connected by an edge if, and only if the corresponding outcomes belong to a common Boolean algebra (measurement); each triangle represents the generators of one of the Boolean algebras.

The probabilities of each triple of outcomes of each measurement should sum to 1, for example, $P(x_4) + P(x_8) + P(y_4) = 1$. There are altogether seven equations of this kind. Combining them with the fact that probability is non-negative it is easy to prove that the probabilities assigned by our rational agent should satisfy $P(x_1) + P(x_8) \leq \frac{3}{2}$ [15]. This is an example of an *uncertainty relation*, a constraint on the probabilities assigned to the outcomes of incompatible measurements. In particular, if the system is prepared in such a way that it is rational to assign $P(x_1) = 1$ then the rules of quantum gambles force $P(x_8) \leq \frac{1}{2}$.

This result is a special case of a more general principle given by [32, 33].

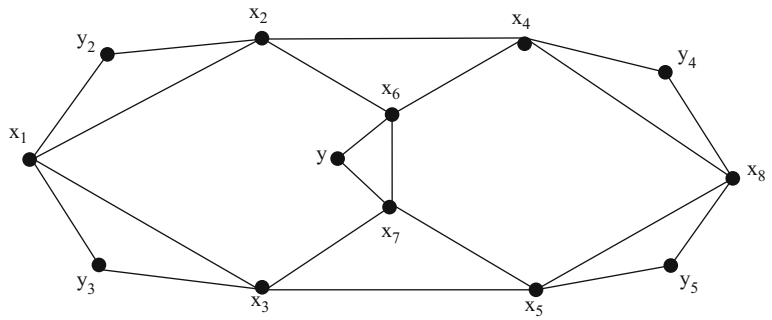

FIGURE 10.2. Cat's cradle.

Theorem 3 *(logical indeterminacy principle) Assuming* $\dim \mathbb{H} \geq 3$, *let x and y be two incompatible atoms in the lattice* $L = L(\mathbb{H})$, *that is,* $x \neq (x \cap y) \cup (x \cap y^{\perp})$. *Then there is a finite set* $\Gamma \subset L(\mathbb{H})$ *with* $x, y \in \Gamma$ *such that every state P on* Γ *satisfies* $P(x) + P(y) < 2$. *In fact we have more:* $P(x), P(y) \in \{0, 1\}$ *if and only if* $P(x) = P(y) = 0$.

This theorem explains the sense in which axiom H4-the axiom of irreducibility-expresses indeterminacy. This axiom asserts that *for every non trivial x there is a y such that* $x \neq (x \cap y) \cup (x \cap y^{\perp})$. By the logical indeterminacy principle the probability value of at least one of the events *x* or *y* must be strictly between zero and one, unless they both have probability zero. Moreover, this fact is already forced by the relation between *x*, *y* and finitely many other events. Remember also that H4 is the only axiom (except SO) that distinguishes between the classical and quantum structures.

3.3 Wondergraph

The previous theorem is typical in the sense that all features of quantum probability, even the quantitative features, can be forced by the logical relations among finitely many events. This follows from a construction of a particular finite set of atoms in $\mathbb{R}^3$ which, together with the orthogonality relations among its elements will be called the *Wondergraph*.

Let us introduce first the notion of *a frame function* which generalizes the concept of a state.

Definition 3 *Let* $\Gamma \subseteq L(\mathbb{H})$ *be a set of atoms of* $L(\mathbb{H})$ *where* $\dim \mathbb{H} = n$. *A frame function on* Γ *is a real function f on* Γ *such that all orthogonal sets of atoms* $x_1, x_2, \ldots, x_n$ *in* Γ *satisfy* $\sum_{j=1}^{n} f(x_j) = C$; *where C is a constant.*

Consider the case of $\mathbb{R}^3$ the smallest space to which Gleason's theorem applies. Let $\overrightarrow{e_1} = (1, 0, 0)$, $\overrightarrow{e_2} = (0, 1, 0)$ and $\overrightarrow{e_3} = (0, 0, 1)$ be the standard basis in $\mathbb{R}^3$ and $\overrightarrow{b_{ij}} = \frac{1}{\sqrt{2}}(\overrightarrow{e_i} + \overrightarrow{e_j})$, $1 \leq i < j \leq 3$. Denote by e_i and b_{ij} the one dimensional subspaces along these vectors. The following theorem turns out to be equivalent to

Gleason's theorem [33]:

Theorem 4 *(Wondergraph theorem) For every atom $z \in L(\mathbb{R}^3)$ there is a finite set of atoms $\Omega(z) \subset L(\mathbb{R}^3)$ such that $e_i, b_{ij}, z \in \Omega(z)$ and such that every frame function f on $\Omega(z)$ which satisfies $f(e_i) = f(b_{ij}) = 0$, and $|f(x)| \leq 1$ for all $x \in \Omega(z)$ necessarily also satisfies $|f(z)| \leq \frac{1}{2}$. Moreover, $|\Omega(z)|$, the number of elements of $\Omega(z)$, is the same for all z.*

Note that the condition $f(e_i) = 0,\ 1 \leq i \leq 3$ for the frame function f on $\Omega(z)$ entails that $f(x) + f(x') + f(x'') = 0$ for all orthogonal triples $x, x', x'' \in \Omega(z)$. To see why Wondergraph theorem entails Gleason's theorem consider first

Lemma 5 *Gleason's theorem for $\mathbb{R}^3$ is true if and only if every bounded frame function f defined on the atoms of $L(\mathbb{R}^3)$ which satisfies $f(e_i) = f(b_{ij}) = 0$ is identically zero.*

The proof of the lemma is straightforward. It follows from the fact that the quadric form $\langle \overrightarrow{x}, A\overrightarrow{x} \rangle$ induced by a self adjoint operator A on $\mathbb{R}^3$ is uniquely determined by the six numbers $\langle \overrightarrow{e_i}, A\overrightarrow{e_i} \rangle$, $\left\langle \overrightarrow{b_{ij}}, A\overrightarrow{b_{ij}} \right\rangle$. Now, to see how Gleason's theorem follows from Wondergraph let f be a *bounded* frame function defined on the atoms of $L(\mathbb{R}^3)$ which satisfies $f(e_i) = f(b_{ij}) = 0$. Normalize f so that $|f(x)| \leq 1$ for all x. Take z to be arbitrary, then the restriction of f to $\Omega(z)$ is a frame function on $\Omega(z)$ and therefore $|f(z)| \leq \frac{1}{2}$. Suppose the atoms of $\Omega(z)$ are $x_1, \ldots, x_s$ and consider the set $\Omega_1(z) = \bigcup_{j=1}^{s} \Omega(x_j)$. The restriction of f to $\Omega_1(z)$ is a frame function on each one of the $\Omega(x_j)$'s. Hence, $|f(x_j)| \leq \frac{1}{2}$ for all $x_j \in \Omega(z)$ and therefore $|f(z)| \leq \frac{1}{4}$. Iterating this process we get that $|f(z)|$ becomes as small as we wish. Since z is arbitrary the theorem follows. Gleason's theorem for any Hilbert space follows from the case of $\mathbb{R}^3$, as Gleason himself showed. Another way to extend the theorem from $\mathbb{R}^3$ to higher real or complex dimensions is to construct Wondergraphs in every (finite) dimension; which can be done once the three dimensional real case is given.

The proof that Gleason's theorem entails the existence of Wondergraph is based on model theory. As a part of the proof one also concludes that there is a known algorithm to construct Wondergraph. The setback is that this algorithm runs very slowly (it is, in fact, the decision algorithm for the theory of real closed fields, which in the worst case runs in doubly exponential time). Thus we pose a

Problem 1 *Construct Wondergraph explicitly.*

Wondergraph allows one to reduce all the interesting quantum phenomena to relations among finitely many events. This follows from:

Corollary 6 *Given a finite set of atomic events Γ_0 and a real number $\varepsilon > 0$ there is a finite set of atoms Γ such that*

a. $\Gamma_0 \subset \Gamma$, *the number of elements* $|\Gamma|$ *of* Γ *depends on* ε *and on* $|\Gamma_0|$ *but not on the elements of* Γ_0.
b. If P *is a state on* Γ *then there is a quantum state* W *(non negative Hermitian operator with trace* 1*) such that*

$$\left| P(x) - < \vec{x}, W\vec{x} > \right| < \varepsilon \quad \textit{for all } x \in \Gamma_0$$

c. There is an algorithm to generate Γ *given* Γ_0 *and* ε.

For many of the famous "paradoxes" of quantum mechanics explicit constructions of the required finite set Γ exist [15, 32, 33]. These include the EPR-Bell argument, the Kochen and Specker theorem, and also generalizations of Kochen and Specker to any given finite number of colors.

On a more fundamental level the importance of these results lies in the way probabilities are associated with L, the algebra of all the possible outcomes of all possible measurements. Remember that in the epistemic conception of probability a "fundamentally sound" method of measuring a person's belief is "by proposing a bet and seeing what are the lowest odds he will accept" . In order to fit the infinite structure L into this view of probability (or any other of the standard Bayesian accounts) we consider only finite segments of L and the probability functions definable on them. These are the quantum gambles considered above. They are the equivalents of classical gambles with dice, roulettes and cards. Some real experiments involve arrangements which are like our gambles: A laboratory device is prepared in such a way that it can perform either one of a few incompatible measurements. Then, the experiment which is actually performed is chosen at random. This gives quantum probability an "operational" flavour and, hopefully removes some of the mystery connected with it, typically expressed by words like "interference" and "superposition".

Another way to see this point is to think about the classical propositional calculus. The Lindenbaum-Tarski algebra on countably many generators gives us all the expressive power we need as far as the propositional connectives are concerned. However, in practice we interpret (assign truth values) only to finite subsets. By analogy, if we take L as representing a "syntax" encompassing symbols for all possible outcomes of all possible measurements, then the "semantics" is the assignment of probability values to finite sections of L. Gleason's theorem, in its Wondergraph version, implies that this "semantics" is, in fact, complete:

Corollary 7 *(Completeness) Suppose that an agent assigns probability values* $P(x)$ *to the elements* x *of a finite* $\Gamma_0 \subset L$*, in a way that contradicts all possible quantum assignments. Then there is a finite* $\Gamma \supset \Gamma_0$*, such that* P *cannot be extended from* Γ_0 *to* Γ. *Hence, in a larger gamble the agent can be shown to be irrational.*

3.4 Solèr's axiom revisited

Let us return to our axiomatic system and the axiom that closes the gap. Recall that Solèr's axiom asserts that for every pair of orthogonal atoms x and y there is

another atom z in the plane they span, which bisects the angle between x and y. More formally: $\mathcal{H}(z;x,y)$, the harmonic conjugate of z with respect to x and y, is orthogonal to z.

Assume that L is infinite dimensional. In this case Solèr's theorem, when coupled with Gleason's theorem, implies that for any (globally defined) state P on L and atoms $x, y \in L$, if $P(x) = 1$ and $P(y) = 0$ then necessarily $x \perp y$, and there is an atom $z \leq x \cup y$ such that $P(z) = \frac{1}{2}$. In other words, there is a precise interpolation between probabilities zero and one.

The axiomatic systems of Bayesian probability theory typically include axioms which imply interpolation of probability values. The most famous (or infamous) one is Ramsey's axiom on the existence of an "ethically neutral" proposition whose probability is one half (axiom 1 in Ramsey's system [4]). The axiom allows Ramsey to construct his theory of utilities (or "values", in his terminology). Savage [5], who wanted to avoid notions like "ethical neutrality", nevertheless also needs an interpolation principle for probabilities, and assumes the existence of arbitrarily refined partitions. This implies that one can obtain propositions with probabilities arbitrarily close to any rational in the interval $[0, 1]$.

I propose to read Solèr's axiom as a probability interpolation axiom; or at any rate to reformulate or replace it by a direct axiom about probabilities. This, however, cannot be straightforward. We are not even guaranteed that a globally defined state exists on L in the first place. However, we can use the fact that certain *finite* orthogonality graphs such as Γ of theorem 3 force any state defined on them to interpolate probability values between zero and one. This is our *logical indeterminacy principle* which expresses probabilistically the basic principle that differentiates the quantum event structure from the classical one. Now, we can turn the tables and assert axiomatically that orthogonality relations like those in Γ are realizable in L. This assertion indirectly expresses the indeterminacy relations in their probabilistic sense. Here, for example, is how this can be done:

Consider $L(\mathbb{R}^3)$ and the rays x, z through the vectors: $\overrightarrow{x} = (1, 0, 0)$, and $\overrightarrow{z} = (1, 1, 0)$ respectively. Let $\Gamma = \Gamma(x, z) \subset L(\mathbb{R}^3)$ be the finite subset of rays guaranteed in theorem 3 (and explicitly constructed in [31, 32]). This means that if P is a state on Γ with $P(x) = 1$ then $0 < P(z) < 1$. Now, consider the rays in Γ and their orthogonality relations *abstractly*, that is, as a graph, which we shall also denote Γ. A candidate to replace Solèr's axiom can then be formulated as :

SO* Let $x, y, x' \in L$ be three orthogonal atoms then there is $z \leq x \cup y$, such that the graph $\Gamma(x, z)$ is realizable in $x \cup y \cup x'$.

There is a way to construct the graph Γ which will make SO* obviously stronger than the original SO. To do this simply add to Γ the rays (and orthogonality relations) which force the relation $\mathcal{H}(z;x,y) \perp z$. In the notations of section 2.4, this means adding rays u, v, s, t and also the rays which, in the space $x \cup y \cup x'$, are orthogonal to the planes $x \cup u$, $z \cup u$, $y \cup v$, $x \cup s$, $z \cup v$, $u \cup t$. But this is cheating, all it shows is that there is a finite graph that forces Solèr's axiom simultaneously with uncertainty. In order to make the axiom more acceptable one has to solve

Problem 2 *Find the minimal* Γ *that forces logical indeterminacy and that allows the proof of Solèr's theorem (and even, perhaps, improves it to include finite dimensional cases).*

Another possible candidate—analogous to Savage's axiom on the existence of arbitrarily fine partitions—is the following:

SO** Let $z \in L$; then the Wondergraph $\Omega(z)$ is realizable in any three dimensional subspace of L that includes z.

The restriction of the graphs we have used to those realizable in $\mathbb{R}^3$ is not essential. It may very well be that a more natural candidate for our Γ or Ω exists, e.g., in $\mathbb{C}^4$.

4 PROBABILITY: RANGE AND CLASSICAL LIMIT

We turn now to the explanatory power of our analysis. The "logic of partial belief" provides straightforward probabilistic, or even combinatorial derivations of a variety of phenomena for which alternative approaches require complicated ad-hoc dynamical explanations. We shall consider two central examples: the first is the EPR paradox and the violation of Bell inequality, and the second is the measurement problem. In particular, we shall discuss the way macroscopic objects can be handled in this framework.

4.1 Bell inequalities

The phenomenological difference between classical and quantum probability is most dramatic when quantum correlations associated with entangled states are concerned. Let us recall what the classical probabilistic analysis of the situation is: A pair of objects is sent from the source, one in Alice's direction, one in Bob's direction. Alice can perform either one of two measurements on her object; she can decide to detect the event x_1 or its absence (which means detecting the event $x_1^{\perp}$). Alternatively, she can decide to check the event x_2 or $x_2^{\perp}$. So each of these measurements has two possible outcomes. Similarly, Bob can test for y_1 or use a different test to detect y_2. Assuming *nothing* about the physics of the situation, and just considering the outcomes we get the following possible logical combinations expressed in the truth table:

x_1	x_2	y_1	y_2	$x_1 \cap y_1$	$x_1 \cap y_2$	$x_2 \cap y_1$	$x_2 \cap y_2$
0	0	0	0	0	0	0	0
...	...	...	...	...	...	...	...
1	1	0	1	0	1	0	1
...	...	...	...	...	...	...	...
1	1	1	1	1	1	1	1

It is the truth table of four propositional variables x_1, x_2, y_1, y_2 and four (out of the six) pair conjunctions, so it has 16 rows, three of them shown explicitly. Each row represents a possible state of affairs regarding the possible outcomes where 1 indicates that the event occurs. Now, suppose that we were to bet on the outcomes.

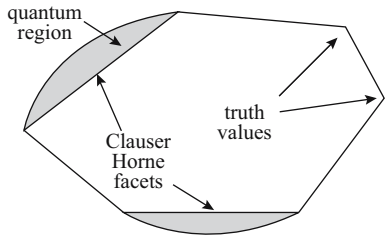

FIGURE 10.3. Correlation polytope.

There are, of course many ways to do this, but they all have to conform with the canons of rationality. The only constraint here is that each one of the 16 possibilities will be assigned a non-negative probability, and the sum of these probabilities be 1. To give this fact a geometric interpretation consider each one of the 16 rows in the truth table as a vector in an 8 dimensional real space, then the vector of probabilities (writing x_iy_j for $x_i \cap y_j$)

$$P = (P(x_1), P(x_2), P(y_1), P(y_2), P(x_1y_1), P(x_1y_2), P(x_2y_1), P(x_2y_2))$$

lies in the convex hull of these 16 vectors, which is a correlation polytope in $\mathbb{R}^8$ with the 16 truth values as vertices shown schematically in Figure 10.3.

The facets of the polytope, are given by linear inequalities in the probabilities, in this case the non-trivial inequalities have the form

$$-1 \leq P(x_1y_1) + P(x_1y_2) + P(x_2y_2) - P(x_2y_1) - P(x_1) - P(y_2) \leq 0 \tag{4}$$

They are called Clauser-Horne inequalities[6], they are among what is generally known as Bell inequalities. Remarkably, in the mid nineteenth century George Boole considered the most general form of the constraints on the values of probabilities of events that can be derived from the logical relations among them. He proved that these constraints have the form of linear inequalities in the probabilities. Paraphrasing Kant he called such constraints *Conditions of Possible Experience*[7].

So far we have been concentrating on the classical picture. What is the quantum mechanical analysis? Again, we shall make no physical assumptions beyond those which are given by the axioms of the event structure. With the two particles we associate a Hilbert space of the form $H \otimes H$, where in case the objects are spin-$\frac{1}{2}$ particles, $\dim H = 2$. The relevant lattice is thus $L = L(H \otimes H)$. The element of L corresponding to the event x_1 is a two dimensional subspace of the form $a_1 \otimes 1$ where $a_1 \in L(H)$ and 1 is the unit in $L(H)$. Similarly, the event corresponding to the outcome y_1 on Bob's side is $1 \otimes b_1$, and likewise for the other cases. The event corresponding to the measurement of x_1 on Alice's side and y_1 on Bob's side is just the intersection:

$$(a_1 \otimes 1) \cap (1 \otimes b_1) = a_1 \otimes b_1$$

Note also that $a_i \otimes 1$, and $1 \otimes b_j$ are compatible. Now, to the eight outcomes

$$a_1 \otimes 1, \quad a_2 \otimes 1, \quad 1 \otimes b_1, \quad 1 \otimes b_2, \quad a_1 \otimes b_1, \quad a_1 \otimes b_2, \quad a_2 \otimes b_1, \quad a_2 \otimes b_2,$$

correspond an 8 dimensional vectors of probability values

$$\begin{aligned}\mathbf{P} = (&P(a_1 \otimes 1), P(a_2 \otimes 1), P(1 \otimes b_1), P(1 \otimes b_2), P(a_1 \otimes b_1),\\ &P(a_1 \otimes b_2), P(a_2 \otimes b_1), P(a_2 \otimes b_2)),\end{aligned}$$

where P is any probability assignment to the elements of $L(H \otimes H)$. When we vary P and the subspaces a_i, b_j, we see that the quantum range is larger than the classical one, and some points lie outside the classical polytope (Figure 10.3), that is, they violate one of the facet inequalities of Clauser and Horne.

From the point of view developed so far this consequence is natural and follows from the event structure of quantum mechanics via Gleason's theorem. We also know from corollary 10 that a violation of a Clauser-Horne inequality can already be depicted in a finite gamble (an explicit construction can be found in [15]). Altogether, in our approach there is no problem with locality and the analysis remains intact no matter what the kinematic or the dynamic situation is; the violation of the inequality is a purely probabilistic effect. Notice that we are just using the quantum event space notion of intersection between (compatible) outcomes: $(a_1 \otimes 1) \cap (1 \otimes b_1) = a_1 \otimes b_1$, as we have used the intersection in the classical event space. The derivation of Clauser-Horne inequalities, indeed of many of Boole's conditions, is blocked since it is based on the Boolean view of probabilities as weighted averages of truth values. This, in turn, involves the metaphysical assumption that there is, simultaneously, a matter of fact concerning the truth values of incompatible propositions such as $x_1 = a_1 \otimes 1$ and $x_2 = a_2 \otimes 1$.

Recall that in section 2.2 we restricted "matters of fact" to include only observable records. Our notion of "fact" is analytically related to that of "event" in the sense that a bet can be placed on x_1 only if its occurrence, or failure to occur, can be unambiguously recorded. However, this leaves open a metaphysical question: Given that x_1 occurred, what is the status of x_2 for which no observable record can exist? Our axioms are not designed to rule out the possibility that x_2 *has* a truth value which we do not know. Initially our approach was agnostic with respect to facts which leave no trace. However, as the above analysis shows, assigning truth values to x_2 and x_1 simultaneously is untenable. In other words, it is prohibited by the axioms *a posteriori*.[8] I believe that Bohr deserves the credit for this insight, although his arguments fall short of establishing it.

We should also recall that there are alternatives to quantum mechanics in which the violations of the Clauser-Horne inequalities have non-local dynamical origins. However, from our perspective the commotion about locality can only come from one who sincerely believes that Boole's conditions are really conditions of possible experience. Since these conditions are just properties of the classical intersection of events, their violation must indicate that something is not kosher with the *measurements*, that is, the choice of a measurement on one side may be correlated with the

outcome on the other. But if one accepts that one is simply dealing with a different notion of probability, then all space-time considerations become irrelevant.

4.2 The BIG measurement problem, the small one, and the classical limit

There are two "measurement problems" The BIG problem, which is illusory, and the small problem which is real and concerns the quantum mechanics of macroscopic systems. The BIG problem concerns those who believe that the quantum state is a real physical state which obeys Schrödinger's equation in all circumstances. In this picture a physical state in which my desk is in a superposition of being in Chicago and in Jerusalem is a real possibility; and similarly a superposed alive-dead cat. In fact the linearity of Schrödinger's equation implies that (decoherence notwithstanding) it is easy to produce states of macroscopic objects in superposition, which seems to contradict our experience, and sometimes, as in the cat case, does not even make much sense.

In our scheme quantum states are just assignments of probabilities to possible events, that is, possible measurement outcomes. This means that the updating of the probabilities during a measurement follows the Von Neumann-Lüders projection postulate and not Schrödinger's dynamics. Indeed, the projection postulate is just the formula for conditional probability that follows from Gleason's theorem. So the BIG measurement problem does not arise. In particular, the cat in the Schrödinger thought experiment is not superposed, but is rather cast in the unlikely role of a particle spin detector. Schrödinger's equation governs the dynamics between measurements; it dictates the way probability assignments should change over time in the absence of a measurement. The general shape of the Schrödinger's equation is not a mystery either; the unitarity of the dynamics follows from the structure of $L(\mathbb{H})$ via a theorem of Wigner [39], in its lattice theoretic form [40]. However, these remarks do not completely eliminate the measurement problem because in our scheme quantum mechanics is also applicable to macroscopic objects.

So suppose that x is one of the rays in the cat's Hilbert space corresponding to a living cat. Let y be one of the atoms corresponding to a dead cat so that $x \perp y$. By Solèr's axiom there is an atom $z \leq x \cup y$ which bisects the angle between x and y. Does this mean that we are back with the BIG measurement problem? The answer is 'No'; remember that z is not a state of the system, it is a possible measurement outcome. It is a mistake to think that by merely following Schrödinger's experiment we are "observing" the event z, or something like it. Obviously we are not, we either see an x-like event, a live cat, or a y-like dead cat event. In order to "see" z we have to devise and perform a measurement such that z is one of its eigenspaces. For reasons that will be explained below, with all probability this is impossible.

But even agreeing that *performing* such a measurement is impossible, we can surely think about operators for which z is an eigenspace, say the projection on z. So let us imagine what one will see when one performs this measurement; what does the event z look like? Presumably, the imagined measuring device is a huge piece of very complicated equipment, because in all likelihood the measurement of the projection

on z involves manipulating individual cat particles. In the end, however, there is a dial with two possible readings 0 and 1, and z is just the event that the dial reads 1. By Lüders' rule the state of the cat after the measurement—assuming that z was the outcome—is the projection on z. The quantum state is not a physical object, it is a representation of our state of knowledge, or belief. The projection on z represents an extremely complex assignment of probabilities to all possible events in a Hilbert space of $\backsim 10^{25}$ particles, an intractable business. One thing is clear, though, there is complete uncertainty about the cat being dead or alive $P(x) = P(y) = \frac{1}{2}$, and of course $P(x \cup y) = 1$.

Ignorance aside, is it not the case that now, after the measurement, there is a matter of fact about the cat being dead or alive? Well, No! As in all such circumstances we cannot say that there is a fact regarding this matter. It is impossible in principle to obtain a record concerning the cat being alive or dead simultaneously with the z-measurement. There is no fundamental difference between the present case and EPR, meaning that we cannot consistently maintain that the proposition " the cat is alive" has a truth value. But the devil is in the details; there is no way to tell from our completely schematic description what is going on in the laboratory. Consequently, there is no way to tell what is the biological state of the cat. It is only after we have mastered the details of the measuring process that we can understand the exact sense in which no record of x or of y is obtainable.

4.3 The weak entanglement conjecture

The small measurement problem is the question why we do not routinely observe events like z for macroscopic objects. More precisely why is it hard to observe macroscopic entanglement, and what are the conditions in which it might be possible? One answer which is certainly valid is decoherence—meaning that it is extremely hard to isolate large pieces of matter and equipment from environmental noise. Decoherence is a dynamical process and its exact character depends on the physics of the situation. I would like to point a possible more fundamental, purely combinatorial reason which is an outcome of the probabilistic structure: *The entanglement of an average ray in a multiparticle Hilbert space is very weak*. To make this intuition precise we have to quantify entanglement, and define what we mean by "average ray".

To make the discussion simpler we shall concentrate on qbits. So our Hilbert space is composed of n copies of the two dimensional complex Hilbert space $\mathbb{H}_n = \mathbb{C}^2 \otimes \mathbb{C}^2 \otimes \cdots \otimes \mathbb{C}^2$, and $\dim \mathbb{H}_n = 2^n$. An atom $s \in L(\mathbb{H}_n)$ is called *separable* if it has the form $s = x_1 \otimes x_2 \otimes \cdots \otimes x_n$ with $x_i \in L(\mathbb{C}^2)$, otherwise an atom is called *entangled*. Also, we shall call the projections on separable (entangled) rays, separable (entangled, respectively) pure states. We keep the letter s to designate separable atoms, and denote by $S \subset L(\mathbb{H}_n)$ the set of all separable atoms. As usual if $x \in L(\mathbb{H}_n)$ is a ray (atom), we shall denote by $\overrightarrow{x}$ a *unit* vector along it.

Now, suppose that we want to observe an entangled atom x. More precisely, we want to obtain a positive proof that it is indeed entangled. To do this we have to design a measurement that will distinguish the ray x from all the separable atoms $s \in S$. A Hermitian operator that does this always exists, and will be called an *entanglement*

witness for x, or in short, a witness. The normalization of witnesses is a matter of convention and for our purpose we shall use the following:

Definition 4 *An Hermitian operator W on* $\mathbb{H}_n = \mathbb{C}^2 \otimes \mathbb{C}^2 \otimes \cdots \otimes \mathbb{C}^2$ *(n copies) is called an entanglement witness if it satisfies*

$$\sup\{|\langle \overrightarrow{s}, W\overrightarrow{s}\rangle| \ ; \ s \in S\} = 1$$

while

$$\|W\| = \sup\{|\langle \overrightarrow{x}, W\overrightarrow{x}\rangle| \ ; \ x \in L(\mathbb{H}_n)\} > 1.$$

So a witness is an observable whose expectation on every separable state is bounded between -1 and 1, while it has an eigenvalue that is larger than 1 in absolute value. Any one-dimensional eigenspace x corresponding to this eigenvalue is obviously entangled. Denote by $\mathcal{W}_n$ the set of all entanglement witnesses on $\mathbb{H}_n$. One way to estimate *how much* a given $x \in L(\mathbb{H}_n)$ is entangled is to calculate

$$\mathcal{E}(x) = \sup\{|\langle \overrightarrow{x}, W\overrightarrow{x}\rangle| \ ; \ W \in \mathcal{W}_n\} \tag{5}$$

A witness W at which the value $\mathcal{E}(x)$ obtains is the best witness for the entanglement of x. If we allow that every measurement involves errors then the larger $\mathcal{E}(x)$ is, the more likely we are to actually observe it. The good news is that *there are rays* $x \in L(\mathbb{H}_n)$ *such that* $\mathcal{E}(x) = \sqrt{2^n}$. These correspond to the the maximally entangled states, the so-called generalized GHZ states[9]. However, it seems that such rays become more and more rare as n increases. To formulate this intuition precisely, let μ_n be the *normalized* uniform (Lebesgue) measure on the unit sphere of $\mathbb{H}_n$. Then we

Conjecture 1 *There is a universal constant* $C > 0$ *such that*

$$\mu_n\left\{\overrightarrow{x} \ ; \ \mathcal{E}(x) > C\sqrt{n \log n}\right\} \to 0 \quad \text{as } n \to \infty \tag{6}$$

A similar result has been established for a large family of witnesses that for each n contains 2^{2^n} witnesses, and which include those that give the best estimation for the GHZ states [44]; hence the conjecture.

I think the conjecture, if true, concerns our ability to observe macroscopic entanglement. There are two types of macroscopic or mesoscopic rays whose entanglement might be witnessed, and the conjecture concerns the second case:

1. There may be relatively rare cases in which the entanglement witness happens to be a thermodynamic observable, that is, an observable whose measurement does not require manipulation of individual particles but only the observation of some global property of the system. There are some indications that this may be the case for some spin chains and lattices [45].
2. Cases of very strong entanglement, like GHZ, which do require many manipulations of individual particles to be observed; however, the value of $\mathcal{E}(x)$ is large

enough to give significant results that rise above the measurement errors. If we assume that the measurement errors are independent, then the total expected error grows exponentially with the number of particles that are manipulated. So, in general, one expects that only x's for which $\mathcal{E}(x)$ is exponential in the number of manipulated particles could yield a significant outcome. The conjecture proposes that the proportion of such x 's is low.

To sum up: the answer to the question "why don't I see chairs in superposition" is twofold, decoherence surely, but even if we could turn it off, there is the combinatorial possibility that "seeing" something like this is nearly impossible. All this, luckily, does not prevent the existence of exotic macroscopic superpositions that can be recorded.

5 MEASUREMENTS

In this paper, all we have discussed is the Hilbert space formalism. I have argued that it is a new kind of probability theory that is quite devoid of physical content, save perhaps the indeterminacy principle which is built into axiom H4. Within this formal context there is no explication of what a measurement is, only the identification of "observables" as Hermitian operators. In this respect the Hilbert space formalism is really just a syntax which represents the set of all possible outcomes, of all possible measurements. It is analogous to the mathematical concept of a probability space, in which certain subsets are identified as events. However, the mathematical theory of probability itself does not tell us the nature of the connection between these formal creatures and real events in the world.

But even before a connection is made between the formal and physical sense of measurement I think there is an interesting philosophical problem here. Our formalism seems to be consistent: there is a *possible* world where measurements and their outcomes behave in the way described above. This would not have been a serious problem if the classical theory of probability were not conceived as *a priori* in some sense. But the theory of probability is a part of what we take as our theory of inference, hence the term '*logic* of partial belief'. As such it is also a ground for the formation of rational expectations. Therefore, the fact that there is a consistent alternative poses a problem similar to the problem that non-Euclidean geometry raised even before general relativity. What should we make of a world in which Boole's conditions of possible experience are violated for no reason other than the structure of probabilities described here?

What is real in the quantum world? Firstly, there are objects—particles about which the theory speaks—which are identified by a set of parameters that involve no uncertainty, and can be recorded in all circumstances and thus persist through time and context [46]. Among them are the rest mass, electric charge, baryonic number, etc. The other part of quantum reality consists of events, that is, recordings of measurements in a very broad sense of the word. Now, one has to distinguish between measurements on the one hand and interactions between material objects on the other. The latter are best described in the Heisenberg picture: There is a time dependent interaction Hamiltonian $H(t)$ which, like any other observable, defines

at every moment t a set of *possible* outcomes, one of the outcomes would obtain *if* $H(t)$ were measured. If, in addition, we have formed a belief about the state of the system at time $t = 0$ (as a result of a previous measurement, say) we automatically have a probability distribution over the set of all possible outcomes of all possible measurements at each t. So each interaction constrains the set of possible outcomes in a certain specific way, and the question which interactions can actually be executed is an empirical question, to be tested by observing the outcomes and their distributions. Measurements are *not* interactions in this sense; although in the broad description of an experiment there is usually an interaction leading to the measurement.

It is impossible to give a precise definition of all the physical processes that deserve the name *measurement*; just as it is not possible to define the term *event* to which the theory of probability can be applied. Even a non-contextual definition of a singular concrete measurement is hard to provide; in this sense measurement outcomes are events "under a description", as philosophers say. Broadly speaking, a measurement is a process in which a material system M, prepared in a specific way, records some aspect of another system S, a recording that effects a permanent change in M, or at least one that lasts long enough. The outcomes to which we have referred throughout the paper are such recordings. Probably the best way to describe measurements is in informational terms. The information recorded by a measurement is *systematic* in the sense that a repeated conjunction of M and S yields the same set of results, and the frequency distribution over the set of results stabilizes in the long run. Of the same importance is the information that is lost during a measurement, the outcome that we could have obtained if any other measurement M' were performed instead of M [47].

This description is broad enough to include the change that photons imprint on the receptors of the retina; it also includes the change caused by a proton hitting a rock on the dark side of the moon. There is nothing specifically human about measurements, nor does M have to be associated with a macroscopic system. What constitutes a "measuring device" cannot be determined beyond this broad description. However, there is a structure to the set of events. Not only does each and every type of measurement yield a systematic outcome; but also the set of all possible outcomes of all measurements—including those that have been realized by an actual recording—hang together tightly in the structure of $L(\mathbb{H})$. This is the quantum mechanical structure of reality.

ACKNOWLEDGEMENTS

This paper is the outcome of a three lecture series that I gave at the Patrick Suppes Center for the Interdisciplinary Study of Science and Technology, Stanford University. I would like to thank Patrick Suppes, Michael Friedman and Thomas Ryckman for their generous hospitality and the lively discussions. I also want to thank William Demopoulos and Ehud Hrushovski for many conversations on questions of philosophy, logic, and mathematics, and Jeremy Butterfield for his comments and suggstions. The research leading to this paper is supported by the Israel Science Foundation grant number 879/02.

NOTES

[1] This position has been expressed often by Feynman [6, 7] . For more references, and an analysis of this point see [8].

[2] The terminology was introduced in [14]. See also [15], and the criticism by Stairs [16]. In case no commitment is made regarding the lattice of subspaces as an event structure, the non contextuality of probability requires a special justification. For example, in the many worlds interpretation [17, 18].

[3] The strong operational approach of Finkelstein [20], and Putnam [21] regarding the logical connectives is -in the most charitable interpretation- a hidden variables theory in disguise, see [22].

[4] At a later stage von Neumann gave up the atomicity assumption. The reason has to do with the absence of a uniform probability distribution over the closed subspaces of an infinite dimensional Hilbert space. The non-atomic structures that resulted are his famous continuous geometries, see [23].

[5] By Möbius and von Staudt. For the standard geometric construction see [25]. A modern account which stresses the algebraic aspects is in Artin's classic [26].

[6] The inequalities were derived in [34]. The sufficiency of the inequalities is due to Fine [35]. The polyhedral structure, its relation to logic, and its generalizations are discussed in [22, 36].

[7] In [37], see also [8]. The parody of Kant is intended, I think. In his classic *The Laws of Thought* Boole writes: "Now what has been said,…, is equally applicable to many other of the debated points in philosophy; such, for instance, as the external reality of space and time. We have no warrant for resolving these into mere forms of the understanding, though they unquestionably determine the *present* sphere of our knowledge" ([38], page 418, my emphasis). So, in the end the joke is on Boole.

[8] This also follows from the logical indeterminacy principle (theorem 3) or the (weaker) Kochen and Specker's theorem [33].

[9] See Mermin [41]. The witnesses that provide the maximum value have a close relation to the facets of the correlation polytope for this case, see [42, 43].

REFERENCES

[1] Ghirardi, G.C., Rimini, A., and Weber, T. Unified dynamics for microscopic and macroscopic systems. *Physical Review D 34*, 470 (1984).

[2] Bohm, D. and Hiley, B.J. *The Undivided Universe: An ontological interpretation of quantum theory*. London: Routledge (1993).

[3] Bell, J. S. *Speakable and Unspeakable in Quantum Mechanics* Cambridge: Cambridge University Press.

[4] Ramsey, F.P. Truth and Probability. (1926) reprinted In D. H. Mellor (ed) *F. P. Ramsey: Philosophical Papers*. Cambridge: Cambridge University Press (1990).

[5] Savage, L.J. *The Foundations of Statistics*. London: John Wiley and Sons (1954).

[6] Feynman, R. P. The concept of probability in quantum mechanics. *Second Berkeley Symposium on Mathematical Statistics and Probability, 1950*, Berkeley: University of California Press, 553 (1951).

[7] Feynman, R. P. and Hibbs, A. R. *Quantum Mechanics and Path Integrals* New York: McGraw-Hill (1965).

[8] Pitowsky , I. George Boole's "conditions of possible experience" and the quantum puzzle. *British Journal for the Philosophy of Science 45*, 95–125 (1994).

[9] Weinberg, S. *Gravitation and cosmology: Principles and Applications of the General Theory of Relativity*, New York: John Wiley & Sons (1972).

[10] Pitowsky, I. Unified field theory and the conventionality of geometry. *Philosophy of Science 51*, 685–689 (1984).

[11] Ben Menahem *Conventionalism*, Cambridge: Cambridge University Press (2005).

[12] Bub, J. *The Interpretation of Quantum Mechanics.* Dordrecht: Reidel (1974).

[13] Bub, J. Quantum Mechanics is About Quantum Information, *Foundations of Physics 35*, 541 (2005).

[14] Barnum, H. Caves, C. M. Finkelstein, J. Fuchs, C. A., and Schack, R. Quantum probability from decision theory? *Proceedings of the Royal Society of London A 456*, 1175 (2000).

[15] Pitowsky, I. Betting on the outcomes of measurements: A Bayesian theory of quantum probability *Studies in the History and Philosophy of Modern Physics 34*, 395 (2003).
[16] Stairs, A. Kipske, Tupman and quantum logic: the quantum logician's conundrum. This volume.
[17] Deutsch, D. Quantum theory of probability and decisions, *Proceedings of the Royal Society of London A455*, 3129 (1999).
[18] Wallace, D. Everettian Rationality: defending Deutsch's approach to probability in the Everett interpretation. *Studies in the History and Philosophy of Modern Physics 34*, 415 (2003).
[19] Birkhoff, G. and von Neumann, J. The logic of quantum mechanics. *Annals of Mathematics 37,* 823 (1936).
[20] Finkelstein, D. Logic of quantum physics. *Transactions of the New York Academy of Science 25*, 621 (1963).
[21] Putnam. H., The logic of quantum mechanics. (1968) reprinted inMathematics Matter and Method — Philosophical Papers Volume1. Cambridge: Cambridge University Press (1975).
[22] Pitowsky, I. *Quantum Probability, Quantum Logic, Lecture Notes in Physics 321*, Heidelberg: Springer (1989).
[23] Rédei, M. Why John von Neumann did not like the Hilbert space formalism of quantum mechanics (and what he liked instead). *Studies in the History and Philosophy of Modern Physics 27*, 493 (1996).
[24] Solovay, R. M. A model of set-theory in which every set of reals is Lebesgue measurable. *Annals of Mathematics 92*, 1 (1970).
[25] Young, W. J. *Projective Geometry* Chicago: Open Court (1930).
[26] Artin, E. *Geometric Algebra* New York: John Wiley & Sons (1957).
[27] Solèr, M. P. Characterization of Hilbert spaces with orthomodular spaces *Communications in Algebra 23*, 219 (1995).
[28] Holland, S. S. Orthomodularity in infinite dimensions; a theorem of M. Solèr. *Bulletin of the American Mathematical Society 32*, 205 (1995).
[29] Gleason, A. M. Measures on the closed subspaces of a Hilbert space. *Journal of Mathematics and Mechanics 6,* 885–893 (1957).
[30] Gödel, K. What is Cantor's continuum problem? in Feferman, S. (ed.) *Kurt Gödel's collected Papers, Vol II* Oxford: Oxford University Press (1990).
[31] Kochen, S. and Specker, E. P. The problem of hidden variables in muantum Mechanics. *Journal of Mathematics and Mechanics 17*, 59–87 (1967).
[32] Pitowsky, I. Infinite and finite Gleason's theorems and the logic of indeterminacy. *Journal of Mathematical Physics 39*, 218 (1998).
[33] Hrushovski, E. and Pitowsky, I. Generalizations of Kochen and Specker's Theorem and the Effectiveness of Gleason's Theorem. *Studies in the History and Philosophy of Modern Physics 35*, 177 (2004).
[34] Clauser, J.F., Horne, M. A., Shimony, A., and Holt, R. A. Proposed experiment to test local hidden-variable theories. *Physical Review Letters 23*, 880 (1969).
[35] Fine, A. Hidden variables, joint probability and Bell inequalities. *Physical Review Letters 48*, 291 (1982).
[36] Pitowsky, I. and Svozil, K. New Optimal tests of quantum nonlocality. *Physical Review A 64*, 4102 (2001).
[37] Boole, G. On the theory of probabilities. *Philosophical Transactions of the Royal Society of London 152*, 225 (1862).
[38] Boole, G. *The laws of Thought* New York: Dover, 1958 (first published in 1854).
[39] Wigner, E. P. *Group Theory and its Applications to Quantum Mechanics of Atomic Spectra*. New York: Academic Press (1959).
[40] Uhlhorn, U. Representation of symmetry transformations in quantum mechanics. *Arkiv Fysik 23*, 307 (1963).
[41] Mermin, N. D. Extreme quantum entanglement in a superposition of macroscopically distinct states *Physical Review Letters*. *65*, 1838 (1990).

[42] Werner, R. F. and Wolf, M. All multipartite Bell correlation inequalities for two dichotomic observables per site. *Physical Review A 64*, 032112 (2001).
[43] Zukowski M. and Brukner C. Bell's theorem for general N-qubit states. *Physical Review. Letters 88*, 210401 (2002).
[44] Pitowsky, I. Macroscopic objects in quantum mechanics-A combinatorial approach. *Physical Review A 70*, 022103 (2004).
[45] Brukner, C. and Vedral, V. Macroscopic thermodynamical witnesses of quantum entanglement *quant-ph/0406040* (2004).
[46] Ben Menahem, Y. Realism and quantum mechanics in A. van der Merve (ed): *Microphysical Reality and Quantum Formalism* Dordrecht: Kluwer, (1988).
[47] Demopoulos, W. Elementary propositions and essentially incomplete knowledge: A framework for the interpretation of quantum mechanics *Noûs 38*, 86 (2004).

MIKLÓS RÉDEI*

11. JOHN VON NEUMANN ON QUANTUM CORRELATIONS

ABSTRACT

In an (unpublished) letter by von Neumann to Schrödinger (dated April 11, 1936) von Neumann replies to Schrödinger's two famous 1935 papers, in which the notion of entanglement between spatially separated quantum systems is introduced and the probabilistic correlations arising from entanglement is discussed from the perspective of a possible clash between quantum mechanics and the principle of physical locality. By quoting extensively from von Neumann's letter it will be seen that von Neumann position concerning such correlations is that they are unproblematic as long as (i) one can (at least in principle) assume that the correlations are explainable by common causes, or (ii) probabilities are interpreted subjectively. It will be argued that while a subjective interpretation of quantum probabilities is difficult to accept in a quantum context, a common cause type explanation of quantum correlations might be possible under a suitable specification of common cause.

1 THE HISTORICAL CONTEXT

In 1935 Einstein, B. Podolsky and N. Rosen published the famous "EPR paper" [1]. The paper's aim was to prove that one should consider the quantum mechanical description of physical reality incomplete – provided that one accepts the *principle of locality*: that the physical state of a subsystem $\mathcal{S}_1$ of a joint system $(\mathcal{S}_1 + \mathcal{S}_2)$ cannot be changed instantaneously by performing a measurement on subsystem $\mathcal{S}_2$ spatially separated from subsystem $\mathcal{S}_1$. The discussions between Einstein, Rosen and Podolsky that led to the EPR paper were taking place in Einstein's office at the Institute for Advanced Study in Princeton in the spring of 1935 [2]. Von Neumann was Einstein's colleague at the Institute for Advanced Study and, given that the EPR paper concerned the completeness of quantum mechanics, in which von Neumann was very much interested (it is well known that he discussed completeness of quantum mechanics at length in his book [3], concluding, on the basis of the famous "no-hidden variable proof", that quantum mechanics is complete), one would expect that Einstein and von Neumann exchanged ideas on this issue. Surprisingly, this does not seem to be the case; to be more precise: the only record I know of that indicates an exchange between von Neuman and Einstein on this subject is von Neumann's unpublished letter to Schrödinger (April 11, 1936). This hand written letter is in the Archive for

* Department of History and Philosophy of Science, Loránd Eötvös University, P.O. Box 32, H-1518 Budapest 112, Hungary, e-mail: redei@ludens.elte.hu

W. Demopoulos and I. Pitowsky (eds.), Physical Theory and its Interpretation, 241–252.

the History of Quantum Mechanics and it will be published in full in [4]. The letter starts with this sentence:

> Einstein has kindly shown me your letter as well as a copy of the Pr.Cambr.Phil.Soc. manuscript. I feel rather more over-quoted than under-quoted and I feel that my merits in the subject are over-emphasized. (Von Neumann to Schrödinger, April 11, 1936), [5]

So Einstein did talk to von Neumann about quantum mechanics after all and tried to raise von Neumann's interest in Schrödinger's paper: "Probability relations between separated systems", Proceedings of the Cambridge Philosophical Society, **32** (1936) 446–452, [6]. This paper is the second of the two famous papers Schrödinger wrote in 1935 [7] and [6]. We know that Einstein and Schrödinger corresponded about the EPR paper in the summer of 1935. Schrödinger's above mentioned two papers were motivated by the EPR paper and by his correspondence with Einstein. Von Neumann's letter to Schrödinger (April 11, 1936) is a direct reply to Schrödinger's second paper [6].

2 SUMMARY OF SCHRÖDINGER'S TREATMENT AND INTERPRETATION OF ENTANGLEMENT

Schrödinger argues in both [7] and [6] that the presence of correlations between spatially separated quantum subsystems of a joint quantum system threatens the principle of locality and thereby might be in contradiction with the theory of relativity. Specifically, Schrödinger considers a composite quantum system described by the tensor product Hilbert space $\mathcal{H}_1 \otimes \mathcal{H}_2$, where $\mathcal{H}_1$ and $\mathcal{H}_2$ are assumed to be identical copies of an L^2 function space describing system $\mathcal{S}_1$ and $\mathcal{S}_2$, respectively. Schrödiner shows in [7] that any state vector $\Psi(x,y) \in \mathcal{H}_1 \otimes \mathcal{H}_2$ of the composite system can be written as

$$\Psi(x,y) = \sum_k a_k g_k(x) \otimes f_k(y) \qquad g_k \in \mathcal{H}_1 \quad f_k \in \mathcal{H}_2, \quad a_k \in \mathbb{C} \qquad (1)$$

with $\{g_k\}$ and $\{f_k\}$ being complete sets of orthogonal (unit) vectors in the respective spaces (not all a_k necessarily nonzero). The decomposition (1) is called the biorthogonal decomposition and it is unique (up to re-labelling of the elements g_k and f_k respectively).

Vectors f_k and g_k can be viewed as eigenvectors (with eigenvalues λ_k^F, λ_k^G) of some observables F and G of system $\mathcal{S}_1$ and $\mathcal{S}_2$, respectively. If we carry out a measurement of G on $\mathcal{S}_2$ and find eigenvalue λ_k^G then "…we have to assign to the *first* system the wave function $g_k(x)$." [6][p. 450]. From the perspective of system $\mathcal{S}_1$, the state of the joint system $(\mathcal{S}_1 + \mathcal{S}_2)$ given by $\Psi(x,y)$ differs from its state given by g_k because the state of $\mathcal{S}_1$ given by g_k is a pure state on $\mathcal{S}_1$, whereas the state given

by vector $\Psi(x,y)$ is a mixed state given by the density matrix

$$\rho_1 = \sum_k |a_k|^2 P_{g_k} \tag{2}$$

here P_{g_k} denotes the one dimensional projection in $\mathcal{H}_1$ that projects to the one dimensional subspace spanned by element g_k and Tr_1 is the trace in $\mathcal{H}_1$. The range $rng(\rho_1)$ of the density matrix ρ_1 is spanned by those g_k for which $a_k \neq 0$.

Schrödinger finds such instantaneous change in system $\mathcal{S}_1$'s state as a result of measurement on the spatially distant system $\mathcal{S}_2$ already troublesome enough; yet, in [6] he goes even further by showing that if h_i is another set of orthogonal unit vectors in $\mathcal{H}_2$ corresponding to eigenvectors of observable H with eigenvalues λ_i^H, then the state $\Phi(x,y)$ can be re-written as

$$\Phi(x,y) = \sum_i w_i \Big[\sum_k \alpha_{ik} g_k(x) \Big] \otimes h_i(y) \tag{3}$$

with constants w_i and α_{ik} depending on the set $\{h_i\}$ and such that the functions $g_i' = \sum_k \alpha_{ik} g_k(x)$ are normalized (but not orthogonal) and belong to $rng(\rho_1)$. Schrödinger points out that by a suitable choice of h_i *every* unit vector in $rng(\rho_1)$ can be obtained as a g_i'; on the other hand, if one carries out a measurement of H on $\mathcal{S}_2$ and finds eigenvalue λ_i^H then one has to assign state g_i' to system $\mathcal{S}_1$. This means that by choosing an appropriate observable H to measure on system $\mathcal{S}_2$ one can transform the state of system $\mathcal{S}_2$ into *any* state of $\mathcal{S}_1$ that lies in the range of ρ_1 (this transformation occurs with probability $|w_i|^2$). Therefore, if Ψ is such that none of the a_k is equal to zero, and, consequently, $rng(\rho_1) = \mathcal{H}_1$, then

> … *in general* a sophisticated experimenter can, by a suitable device which does *not* involve measuring non-commuting variables, produce a non-vanishing probability of driving the system [$\mathcal{S}_1$] into any state he chooses … [6][p. 446]

> 4. Indubitably the situation described here is, in present quantum mechanics, a necessary and indispensable feature. The question arises, whether it is so in Nature too. I am not satisfied about there being sufficient experimental evidence for that. Years ago I pointed out [Schrödinger's footnote: Annalen der Physik (4), 83 (1927), 961. Collected Papers (Blackie and Son, 1928), p. 141.] that when two systems separate far enough to make it possible to experiment on one of them without interfering with the other, they are bound to pass, during the process of separation, through stages which were beyond the range of quantum mechanics as it stood then. For it seems hard to imagine a complete separation, whilst the systems are still so close to each other, that, from the classical point of view, their interaction could still be described as an unretarded *actio in distans*. And ordinary quantum mechanics, on account of its thoroughly unrelativistic character, really only deals with the *actio in distans* case. The whole

> system (comprising in our case both systems) has to be small enough to be able to neglect the time that light takes to travel across the system, compared with such periods of the system as are essentially involved in the changes that take place.
> Though in the mean time some progress seemed to have been made in the way of coping with this condition (quantum electrodynamics), there now appears to be a strong probability (as P. A. M. Dirac [Schrödinger's footnote: P.A.M. Dirac, Nature. 137 (1936), 298] has recently pointed out on a special occasion) that this progress is futile. [6] [p. 451]

3 VON NEUMANN'S REPLY TO SCHRÖDINGER

In his letter to Schrödinger (April 11, 1936) von Neumann reacts to the above passage in Schrödinger's paper:

> I cannot accept your § 4. completely. I think that the difficulties you hint at are "pseudo-problems". The "action at distance" in the case under consideration says only that even if there is no dynamical interaction between two systems (e.g. because they are far removed from each other), the systems can display statistical correlations. This is not at all specific for quantum mechanics, it happens classically as well.
> (von Neumann to Schrödinger, April 11, 1936) [5]

To illustrate his point, von Neumann gives the following simple example in the letter:

> Let S_1 and S_2 be two boxes. One knows that 1,000,000 years ago either a white ball had been put into each or a black ball had been placed into each but one does not know which color the balls were. Subsequently one of the boxes (S_1) was buried on Earth, the other (S_2) on Sirius. So one has this probability distribution:

	S_1 white	S_1 black
S_2 white	$\frac{1}{2}$	0
S_2 black	0	$\frac{1}{2}$

> Or for S_2 only:

S_2	white	$\frac{1}{2}$
S_2	black	$\frac{1}{2}$

> Now one digs S_1 on Earth out, opens it and sees: the ball is white. This action on Earth changes instantaneously the S_2 statistic on Sirius to

S_2	white	1
S_2	black	0

> (The situation reminds one of the well-known joke: *"Au moment même [Von Neumann's insertion: "Als mit seiner Frau etwas in Paris geschah."] mon Colonel était cocu á Madascar."*
> (von Neumann to Schrödinger, April 11, 1936), [5]

Together with von Neumann's insertion the French sentence reads something like *The moment something happened to his wife in Paris the colonel was cuckolded in Madagascar*.

Von Neumann also reacts to Schrödinger's skeptical remark about the prospects of relativistic quantum field theory:

> And of course quantum electrodynamics proves that quantum mechanics and the special theory of relativity are compatible "philosophically" – quantum electrodynamic fails only because of the concrete form of Maxwell's equations in the vicinity of a charge.
> (von Neumann to Schrödinger, April 11, 1936), [5]

4 COMMENTS ON VON NEUMANN'S REPLY

There are two different ideas in von Neumann's reply to Schrödinger by which von Neumann avoids a potential clash between presence of distant correlations and the principle of "no action at a distance": One is to interpret probabilities subjectively as measures of lack of knowledge. If the probabilities are just measures of our ignorance then no real, physical change takes place when the probability distribution on $\mathcal{S}_2$ changes as a result of manipulation on $\mathcal{S}_2$. (This idea is displayed nicely by the joke von Neumann recalls: the real physical change occurs only in Madagascar; the colonel's getting cuckolded is just a "semantic change".) It is difficult however to accept a subjective interpretation of probability in physics because it is a *fact* that probability statements are verified in quantum mechanics by counting frequencies.

The other idea is that presence of correlations between spatially separated quantum systems is not, in and by itself, reason for concern as long as one has an acceptable explanation of those correlations. In the example the explanation is the action of putting the white balls into the boxes. This action – together with the subsequent observations on Earth and Syrius – can be translated into probabilistic terms in the following way: Let $A, A^{\perp}, B, B^{\perp}$ and $C, C^{\perp}$ be the following events:

A	white ball observed in box upon opening box on Earth
$A^{\perp}$	black ball observed in box upon opening box on Earth
B	white ball observed in box upon opening box on Syrius
$B^{\perp}$	black ball observed in box upon opening box on Syrius
C	placing white balls in both boxes on Earth
$C^{\perp}$	placing black balls in both boxes on Earth

If $p(A) = p(B) = 1/2$ and $p(A^{\perp}) = p(B^{\perp}) = 1/2$ are the initial probabilities of the respective events, then A and B are positively correlated:

$$\frac{1}{2} = p(AB) > p(A)p(B) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4} \tag{4}$$

Furthermore we have

$$p(AB|C) = p(A|C) = p(B|C) = 1 \tag{5}$$

$$p(AB|C^{\perp}) = p(A|C^{\perp}) = p(B|C^{\perp}) = 0 \tag{6}$$

Consequently, the following conditions hold:

$$p(AB|C) = p(A|C)p(B|C) \tag{7}$$

$$p(AB|C^{\perp}) = p(A|C^{\perp})p(B|C^{\perp}) \tag{8}$$

$$p(A|C) > p(A|C^{\perp}) \tag{9}$$

$$p(B|C) > p(B|C^{\perp}) \tag{10}$$

Definition 1 *Given a positive correlation*

$$p(AB) > p(A)p(B) \tag{11}$$

between events A and B in a probability space $(\mathcal{S}, p)$ *with Boolean algebra* $\mathcal{S}$ *and probability measure p on* $\mathcal{S}$*, an event* $C \in \mathcal{S}$ *is called a common cause of the correlation (11) if it satisfies conditions (7)–(10).*

This definition of common cause is due to H. Reichenbach [8], the event C having the properties (7)-(10) is called a (Reichenbachian) common cause. Reichenbach also seems to have embraced what came to be called the *Common Cause Principle*: If events A, B are correlated then the correlation is either due to a causal influence between the correlated events or there exists a common cause of the correlation. (Reichenbach himself never formulated the principle explicitly, this was articulated later, especially by W. Salmon [9]).

The intuitive reason why one does not regard the correlation in von Neumann's example problematic is thus that in the example the act of placing the white balls into the boxes serves as a natural "common cause" of the correlation, and such a common cause "explains" the correlation in the sense of entailing it. Viewed from this perspective, the intended message of von Neumann's example seems to be that in the case of quantum correlations Schrödinger is considering the same should be the case; that is to say, von Neumann seems to take the position that the correlations entailed by entanglement also might have an explanation in terms of (Reichenbachian) common causes.

Problem is that, unlike in von Neumann's simple example, in the case of correlations arising from entanglement, there are no obvious candidates for events that could qualify as common causes. Lack of common cause candidate events has led to the suspicion that such events *cannot* exist at all. Indeed, there exist a number of "proofs" in the literature to the effect that common cause events of correlations entailed by entanglement *cannot* exist: Van Fraassen was the first to link Reichenbach's notion of common cause to the EPR correlations [10] and he concluded that no common cause explanation of EPR correlations is possible. Since Fraassen's work it has become generally accepted that the quantum correlations violate the Common Cause Principle. A careful look at the problem shows however that it is more difficult to rule out common causes of correlations than one would have thought.

Let $(\mathcal{S},p)$ be a classical probability space and $A,B \in \mathcal{S}$ two events that are correlated in p. The Common Cause Principle says that if A and B are *causally independent*, then there has to exist a common cause of the correlation. The crucial observation that makes an attempt to falsify the Common Cause Principle very difficult is that the Principle does *not* require the common cause to be part of the event structure $\mathcal{S}$: it may very well be the case that there is no event C in $\mathcal{S}$ that qualifies as a common cause of the correlation between A and B but this fact in and by itself does *not* entail any violation of the Common Cause Principle, for it may very well be the case that $\mathcal{S}$ is just too small, and there might exist "hidden" common causes of the given correlation – "hidden" in the sense of belonging to an event structure $\mathcal{S}'$ which is larger than $\mathcal{S}$. It turns out that such a defence of the Common Cause Principle is always possible, for one has the following result [11]:

Proposition 1 *Given any classical probability space* $(\mathcal{S},p)$ *and a correlation* $Corr_p(A,B) = p(AB) - p(A)p(B) > 0$ *in it, there exists an extension* $(\mathcal{S}',p')$ *of* $(\mathcal{S},p)$ *such that there exists an event* C *in* $\mathcal{S}'$ *which is the common cause of the correlation* $Corr_p(A,B) > 0$.

(Note that $(\mathcal{S}',p')$ is an extension of $(\mathcal{S},p)$ if there exists a Boolean algebra homomorphism $h\colon \mathcal{S} \to \mathcal{S}'$ such that $p'(h(X)) = p(X)$ for every $X \in \mathcal{S}$.)

Proposition 1 entails that one can only hope to be able to falsify the Common Cause Principle if one requires of the common cause to satisfy some additional conditions that are not part of the definition of common cause. The additional conditions are typically "locality conditions": probabilistic independence conditions intended to express in probabilistic terms consequences of relativistic locality (causality) principles. As can be expected, the question of whether local common cause type explanations of EPR correlations are possible, depends very sensitively on how the locality conditions are formulated: It turns out that under *some* formulations of those additional locality conditions hidden common causes still *cannot* be ruled out [12]. If one formulates more stringent locality conditions, then it is an *open problem* whether such strongly local hidden common causes can exist (see the review [13]). Recently, new locality conditions have been suggested that seem to exclude common cause explanations of

EPR correlations (see [14]), a critical evaluation of these locality conditions is yet to be carried out, however.

Once one starts talking about notions of relativity theory in connection with the discussion of quantum correlations, the proper framework in which one should deal with the problem of distant quantum correlations is relativistic quantum field theory. A specific approach to quantum filed theory is local, algebraic relativistic quantum field theory (ARQFT), axioms of which were worked out during the late fifties (see Haag's book [15] for a comprehensive presentation of the theory). It turned out that correlations between spacelike separated observables is even more endemic in local quantum field theory than in non-relativistic quantum mechanics (see the review paper [16] and [17] for a more recent result). Presence of those spacelike correlations in relativistic quantum field theory raises the problem of the status of the Common Cause Principle in physics in an even more dramatic way because relativistic quantum field theory is a physical theory which, by its very construction, is supposed to be complying with locality and causality principles as these are understood in the spirit of the special theory of relativity. To formulate precisely the question of whether quantum field theory satisfies the Common Cause Principle and to recall the only result known in this connection, we need some definitions:

Let $\mathcal{N}$ be a von Neumann algebra, $\mathcal{P}(\mathcal{N})$ its projection lattice and ϕ a normal state on $\mathcal{N}$. Recall that two elements $A, B \in \mathcal{P}(\mathcal{N})$ are called *compatible* if there exists a distributive sublattice of $\mathcal{P}(\mathcal{N})$ containing both A and B.

Definition 2 *Let $A, B \in \mathcal{P}(\mathcal{N})$ be two compatible elements that are correlated in ϕ:*

$$\phi(A \wedge B) > \phi(A)\phi(B) \tag{12}$$

$C \in \mathcal{P}(\mathcal{N})$ is a common cause of the correlation (12) if C is compatible with both A and B and the following conditions (completely analogous to (7)-(10)) hold

$$\frac{\phi(A \wedge B \wedge C)}{\phi(C)} = \frac{\phi(A \wedge C)}{\phi(C)} \frac{\phi(B \wedge C)}{\phi(C)} \tag{13}$$

$$\frac{\phi(A \wedge B \wedge C^{\perp})}{\phi(C^{\perp})} = \frac{\phi(A \wedge C^{\perp})}{\phi(C^{\perp})} \frac{\phi(B \wedge C^{\perp})}{\phi(C^{\perp})} \tag{14}$$

$$\frac{\phi(A \wedge C)}{\phi(C)} > \frac{\phi(A \wedge C^{\perp})}{\phi(C^{\perp})} \tag{15}$$

$$\frac{\phi(B \wedge C)}{\phi(C)} > \frac{\phi(B \wedge C^{\perp})}{\phi(C^{\perp})} \tag{16}$$

For a point x in the Minkowski space $\mathcal{M}$ let $BLC(x)$ denote the backward light cone of x; furthermore for an arbitrary spacetime region V let $BLC(V) \equiv \cup_{x \in V} BLC(x)$.

For spacelike separated spacetime regions V_1 and V_2 let us define the following regions

$$wpast(V_1, V_2) \equiv (BLC(V_1) \setminus V_1) \cup (BLC(V_2) \setminus V_2) \tag{17}$$

$$cpast(V_1, V_2) \equiv (BLC(V_1) \setminus V_1) \cap (BLC(V_2) \setminus V_2) \tag{18}$$

$$spast(V_1, V_2) \equiv \cap_{x \in V_1 \cup V_2} BLC(x) \tag{19}$$

Obviously it holds that

$$spast(V_1, V_2) \subseteq cpast(V_1, V_2) \subseteq wpast(V_1, V_2) \tag{20}$$

Definition 3 *Let $\{\mathcal{N}(V)\}$ be a net of local von Neumann algebras over Minkowski space satisfying the standard axioms of ARQFT (isotony, Einstein locality, Poincaré covariance, weak additivity, spectrum condition). Let V_1 and V_2 be two spacelike separated spacetime regions, and let ϕ be a locally normal state on the quasilocal algebra $\mathcal{A}$. If for any pair of projections $A \in \mathcal{N}(V_1)$ and $B \in \mathcal{A}(V_2)$ it holds that if*

$$\phi(A \wedge B) > \phi(A)\phi(B) \tag{21}$$

then there exists a projection C in the von Neumann algebra $\mathcal{N}(V)$ which is a common cause of the correlation (21) in the sense of Definition 2, then the local system is said to satisfy

Weak Common Cause Principle: *if* $V \subseteq wpast(V_1, V_2)$

Common Cause Principle: *if* $V \subseteq cpast(V_1, V_2)$

Strong Common Cause Principle: *if* $V \subseteq spast(V_1, V_2)$

We say that Reichenbach's Common Cause Principle holds for the net (respectively holds in the weak or strong sense) iff for every pair of spacelike separated spacetime regions V_1, V_2 and every normal state ϕ, the Common Cause Principle holds for the local system $(\mathcal{N}(V_1), \mathcal{N}(V_2), \phi)$ (respectively in the weak or strong sense).

Problem: Does any of the above Common Cause Principles hold in quantum field theory ?

If V_1 and V_2 are complementary wedges then $spast(V_1, V_2) = \emptyset$. Since the local von Neumann algebras pertaining to complementary wedges are known to contain correlated projections (see [16]), the *Strong* Reichenbach's Common Cause Principle trivially fails in AQFT.

The question of whether Reichenbach's Common Cause Principle holds in AQFT was first formulated in [18] (see also [19]) and the answer to it is not known. What is known is that the *Weak* Reichenbach's Common Cause Principle typically holds

under mild assumptions on the local net $\{\mathcal{N}(V)\}$:

Proposition 2 *If a net $\{\mathcal{N}(V)\}$ with the standard conditions (isotony, Einstein locality, Poincaré covariance, weak additivity, spectrum condition) is such that it also satisfies the local primitive causality condition and the algebras pertaining to double cones are type III, then every local system $(\mathcal{N}(V_1), \mathcal{N}(V_2), \phi)$ with V_1, V_2 contained in a pair of spacelike separated double cones and with a locally normal and locally faithful state ϕ satisfies Weak Reichenbach's Common Cause Principle.*

(See [20] for the proof of the above proposition and for additional analysis of the status of Reichenbach's Common Cause Principle in quantum field theory.)

Local primitive causality is a condition that expresses the hyperbolic character of time evolution in AQFT. For a spacetime region V let $V'' = (V')'$ denote the *causal completion* (also called causal closure and causal hull) of V, where V' is the set of points that are spacelike form every point in V. The net $\{\mathcal{N}(V)\}$ is said to satisfy the *local primitive causality* condition if $\mathcal{N}(V'') = \mathcal{N}(V)$ for every nonempty convex region V.

5 CONCLUDING REMARKS

Von Neumann's letter to Schrödinger is probably the first formulation of Reichenbcah's Common Cause Principle in connection with quantum correlations – but an implicit formulation only since the technically explicit notion of common cause does not appear in the letter. It also is worth pointing out that, apparently, von Neumann did not see a major difference between quantum and classical correlations from the perspective of the Common Cause Principle.

Once however the notion of common cause is specified in the sense of Reichenbach's definition, the concept of common cause and the status of the associated Common Cause Principle can be subjected to a rigorous analysis. The analysis has led to a number of precise problems, some of which are still open. Specifically, it is not known whether relativistic quantum field theory is causally rich enough to be able to give a *local* common cause explanation of the spacelike correlations it predicts. Nor has it been proven in full generality that *local* common causes of standard EPR correlations cannot exist.

The difficulty with proving precise theorems concerning the Common Cause Principle is that the notion of (Reichenbachian) common cause is rather subtle, hence problems relevant for the status of the Principle are non-trivial. It is difficult even to decide whether a given probability space $(\mathcal{S}, p)$ is *causally closed* in the sense that it contains a common cause of every correlation between correlated events A, B that are causally *independent*, $R_{ind}(A, B)$. It turns out that causal closedness of $(\mathcal{S}, p)$ with respect to a causal independence relation R_{ind} depends sensitively on both $(\mathcal{S}, p)$ and R_{ind}, and, while causal closedness is not impossible – it is possible even if the event structure $\mathcal{S}$ contains a finite number of elements [21] – it does not always hold. There does

not seem to exist a canonical procedure by which causal closedness could be verified, and there are a number of open questions concerning causal closedness (see [21]).

ACKNOWLEDGEMENT

I wish to thank Marina von Neumann Whitman for her permission to quote from von Neumann's letter to Schrödinger. Work supported by OTKA (contract number: T 43642) and by Alexander von Humboldt Foundation (through a Sofja Kovalevskaja Award).

REFERENCES

[1] A. Einstein, B. Podolsky, N.Rosen: Can Quantum Mechanical Description of Physical Reality Be considered Complete? *Physical Review* **47** (1935) 777–780

[2] M. Jammer: The EPR Problem in its Historical Development in P. Lahti and P.Mittelstaedt (eds), *Symposium on the Foundations of Modern Physics. 50 years of the Einstein-Podolsky-Rosen Gedankenexperiment* (Singapore: World Scientific, 1985) 129–149

[3] J. von Neumann: *Mathematische Grundlagen der Quantenmechanik* (Dover Publications, New York, 1943) (first American Edition; first edition: Springer Verlag, Heidelberg, 1932; first English translation: Princeton University Press, Princeton, 1955.)

[4] John von Neumann: *Selected Letters*, M. Rédei ed., (American Mathematical Society and London Mathematical Society, 2005).

[5] John von Neumann to Erwin Schrödinger (April 11, 1936), in [4], 211–213

[6] E. Schrödinger: "Probability relations between separated systems" *Proceedings of the Cambridge Philosophical Society* **32** (1936) 446–452

[7] E. Schrödinger: Discussion of probability relations between separated systems *Proceedings of the Cambridge Philosophical Society* **31** (1935) 555–563

[8] H. Reichenbach: *The Direction of Time* (University of California Press, Los Angeles, 1956)

[9] W.C. Salmon: *Scientific Explanation and the Causal Structure of the World* (Princeton University Press, Princeton, 1984)

[10] B.C. Van Fraassen: "The Charybdis of Realism: Epistemological Implications of Bell's Inequality" in J. Cushing and E. McMullin (eds.), *Philosophical Consequences of Quantum Theory* (Notre Dame: University of Notre Dame Press, 1989) 97–113

[11] G. Hofer-Szabó, M. Rédei and L.E. Szabó: On Reichenbach's Common Cause Principle and Reichenbach's notion of common cause *The British Journal for the Philosophy of Science* **50** (1999) 377–399

[12] Szabó, L.E.: Attempt to resolve the EPR-Bell paradox via Reichenbach's concept of common cause *International Journal of Theoretical Physics* **39** (2000) 901–911

[13] M. Rédei: Reichenbach's Common Cause Principle and quantum correlations in *Modality, Probability and Bell's Theorems*, NATO Science Series, II. Vol. 64., T. Placek and J. Butterfield (eds.) (Kluwer Academic Publishers, Dordrecht, Boston, London, 2002) 259–270

[14] G. Grasshoff, S. Portmann and A Wüthrich: Minimal assumption derivation of a Bell-type inequality *The British Journal for the Philosophy of Science* **56** (2005) 663–680

[15] R. Haag: *Local Quantum Physics* (Springer Verlag, Berlin, 1992)

[16] S.J. Summers: On the independence of local algebras in quantum field theory, *Reviews in Mathematical Physics* **2** (1990) 201–247

[17] H. Halvorson and R. Clifton: Generic Bell correlation between arbitrary local algebras in quantum field theory *Journal of Mathematical Physics* **41** (2000) 1711–1717

[18] M. Rédei: Reichenbach's common cause principle and quantum field theory *Foundations of Physics* **27** (1997) 1309–1321
[19] M. Rédei: *Quantum Logic in Algebraic Approach* (Kluwer Academic Publishers, Dordrecht, 1998)
[20] M. Rédei and S.J. Summers: Local primitive causality and the Common Cause Principle in quantum field theory *Foundations of Physics* **32** (2002) 335–355
[21] B. Gyenis and M. Rédei: When can statistical theories be causally closed? *Foundations of Physics* **34** (2004) 1285–1303

ALLEN STAIRS*

12. KRISKE, TUPMAN AND QUANTUM LOGIC: THE QUANTUM LOGICIAN'S CONUNDRUM

ABSTRACT

Almost thirty years ago, Saul Kripke gave a talk in which he offered an extended critique of quantum logic. Neither that talk nor any commentary on it appear in the published literature. Today, there is much less interest in quantum logic as an interpretive program in the foundations of quantum mechanics. Nonetheless, Kripke's critique raises interesting issues about what it might mean to contemplate a change in logic. Set against the larger background of the literature at that time, the lecture also provides an interesting springboard for exploring a number of issues about realism and quantum mechanics of the sort that Jeff Bub has wrestled with over his career. This paper will present an extended summary of a related critique by one P. Kriske, and will proceed from there to a discussion of the larger questions that must be addressed in order to provide an adequate reply to Kriske.

I've known Jeff Bub for over thirty years as a teacher, colleague and friend, and I'm delighted to be able to contribute to this volume in his honor. What I plan to do, however, is start with some unpublished material from a dissertation that I wrote almost 30 years ago and that I had not even held in my hands for almost that long. It's a bit like talking to a ghost. As it turns out, that may be appropriate; ultimately, the problem I want to worry is the peculiarly elusive nature of the attempt to interpret quantum theory. We'll begin, however, with a quasi-mythical episode in the history of quantum logic. The episode has its own interest, but it will also serve as a segue into a broader discussion.

1 QUANTUM LOGIC AND REALIST DREAMS

When I wrote my dissertation in 1978, some of us at Western Ontario saw quantum logic as the key to a realist interpretation of quantum mechanics. One important part of what we meant by "realist" was that in the ideal case, measurement should merely reveal the pre-existing values of physical quantities: if the measuring instrument said that the y-spin was $+1/2$, that was supposed to be because it really had that value before the measurement was made. But since we can perform any measurement we like, that implied that all quantities would have to have simultaneous values – whether we could measure the quantities simultaneously or not.

* Dept. of Philosophy, University of Maryland, College Park, MD 200742 USA; Email: stairs@umd.edu.

W. Demopoulos and I. Pitowsky (eds.), Physical Theory and its Interpretation, 253–272.

Whether or not this was a reasonable understanding of realism, the difficulty is clear: results such as Kochen and Specker's[1] apparently show that the physical quantities couldn't possibly all have values at once. If two quantities Q and Q' share an eigenspace S, the K&S theorem assumes that the values of the quantities are accordingly related: either both have a value that goes with S or neither does. If "$Q = q$" and "$Q' = q'$" represent the same proposition when they are associated with the same subspace, claiming that the finite set of K&S quantities all have values at once amounts to a classical contradiction.

One way to provide for definite values is to reject the K&S constraints and adopt a contextual hidden variable theory. Unfortunately, this comes at a price: what we're measuring "here" must either be influenced or partly constituted by what's being measured "there"; otherwise, we run afoul of Bell's theorem. Quantum logic proposed a way around this problem: identify propositions as the Hilbert space suggests. Since propositions are identified in a context-free way, local quantities are locally constituted. If a quantity Q has possible values $q_1, q_2, \ldots, q_n$ then the quantum logical disjunction

$$Q = q_1 \vee Q = q_2 \vee \cdots \vee Q = q_n$$

is true. The conjunction of all these disjunctions yields a classical contradiction, but logic isn't classical and properties don't mesh as classical logic says they do.[2] That would allow the serpentine Kochen and Specker "contradiction" to be a truth – a truth that supposedly says of each quantity that it takes one of its possible values. If we add the claim that measurement simply reveals pre-existing values, then not only are quantities locally defined, but local measurement results don't depend for their outcomes on distant measurements.

2 INTRODUCING PROFESSORS KRISKE AND TUPMAN

There are two thoughts here. One is that quantum logic allows us to say that all quantities have value, revealed by measurement. The other is that changes in physics might rightly induce us to accept changes in logic. In 1974, Saul Kripke gave a lecture at the University of Pittsburgh in which he offered a critique of quantum logical *value-definiteness*, the thesis that all quantum mechanical quantities have values, which measurement simply reveals. He also called into question the very idea that logic might change in response to empirical discoveries. That lecture was the subject of rumors, myths and much conversation in corridors. It was also the topic of the first two papers of my dissertation. I'd like to discuss what Kripke said, but this presents a problem. He never published his talk, which is why I never published the relevant portions of the dissertation. Worse still, the tape I once possessed is long since lost. All I have are the quotes and paraphrases in the dissertation. There's also the matter of propriety. Since Kripke's talk never appeared in print, it isn't part of the public record of positions he's committed himself to. He might well object to being saddled with what he said back then.

I propose the following solution. I'll discuss a position that wouldn't have occurred to me if I had never heard the tape of Kripke's talk, but I won't promise that the position is Kripke's. I'll attribute it to the fictitious philosopher Paul Kriske – Kriske for short. If I attribute something to Kriske, you may assume that I didn't think of it myself. But you may *not* assume that it's an accurate reflection of Kripke's view. It may, for all I'm willing to claim, be based on a complete misunderstanding of what Kripke actually said. Since my tape no longer exists, and since I have nothing close to a full transcript of it, you should take this possibility very seriously.

The paper that Kriske singles out for his critique is Putnam's 1968 essay "Is Logic Empirical?"[3] But just as fairness led to the introduction of Kriske into our discussion, it's also fair to ask if what this Kriske has to say about Putnam is true to the real-life Putnam. Since Putnam exegesis is not my concern, I will introduce another character into our drama, Prof. Tupman, who is the subject of Kriske's criticism.

3 KRISKE ON TUPMAN

Suppose that A and B are two non-commuting operators, each with eigenvalues 1 and 2. Tupman wants to say that both of these statements are true as ordinarily understood, and before any measurements are made:

(1) $A = 1$ or $A = 2$ (that is, $A = 1 \vee A = 2$)
(2) $B = 1$ or $B = 2$ (that is, $B = 1 \vee B = 2$)

Nonetheless, Tupman also wants to say that all of these are true as well:

(3) $\sim (A = 1 \wedge B = 1)$
(4) $\sim (A = 1 \wedge B = 2)$
(5) $\sim (A = 2 \wedge B = 1)$
(6) $\sim (A = 2 \wedge B = 2)$

You might have thought that for A and B to have values, one of the following would have to be true:

(3′) $A = 1 \wedge B = 1$
(4′) $A = 1 \wedge B = 2$
(5′) $A = 2 \wedge B = 1$
(6′) $A = 2 \wedge B = 2$

Tupman's claim is that so long as (1) and (2) are true, we have all we need for value-definiteness. He frames the issue in terms of the distributive law. In classical logic,

$$(W \vee X) \wedge (Y \vee Z)$$

implies

$$(W \wedge Y) \vee (W \wedge Z) \vee (X \wedge Y) \vee (X \wedge Z)$$

by the distributive law. But the distributive law doesn't hold in the lattice of subspaces of a vector space, and according to Tupman, that lattice reflects the correct logic. Hence, we can't move from the conjunction of (1) and (2) to (3) through (6).

To make this more palatable, Tupman offers an analogy with geometry. Before relativity and non-Euclidean geometry, the idea that two *straight* lines might be a constant distance apart over some interval but intersect further along would be an intuitive contradiction. As it turns out, however, this "contradiction" about space might well be true. The moral? Don't trust intuition – not even in cases where ignoring it feels like a contradiction.

As Kriske reads Tupman, the distributive law amounts to an axiom of classical logic, and is up for grabs once rival systems are on the table. With the right sorts of empirical pressures, we might abandon one formal system for another. In the case of quantum mechanics, Tupman sees this as the smoothest course to follow. Give up the distributive law and adopt quantum logic; the payoff for the intuitive pain is a realist interpretation of quantum theory.

Kriske points out that we *seem* to be able to knock this view down with a simple argument. Tupman says that A and B have values that show up when we make a measurement. Suppose we measure A and find that

(7) $A = 1$.

Tupman says that B has one of the two values 1 or 2; that's what (2) above tells us. But now reason by cases. If B has the value 1, then $A = 1$ and $B = 1$, and that contradicts (3) above; if B has the value 2, then $A = 1$ and $B = 2$, which contradicts (4). Since there aren't any more cases, there's no way for A and B both to have values.

Kriske thinks that's as complete a refutation of Tupman as we could hope for, but he assumes that he'll be accused of begging the question. His refutation of Tupman, so his opponent will say, called on the distributive law, which is what the reasoning by cases amounts to here. But the distributive law is exactly what's at issue.

Kriske's reply reveals the heart of his position. His refutation depended on reasoning from

(7) $A = 1$

and

(2) $B = 1$ or $B = 2$

to the conclusion

C: $(A = 1$ and $B = 1)$ or $(A = 1$ and $B = 2)$.

However, that conclusion is just what Tupman rejects when he says that

(3) $\sim (A = 1 \wedge B = 1)$

and

(4) $\sim (A = 1 \wedge B = 2)$

are both true. Apparently, then, Tupman doesn't think that C follows from (7) and (2). Apparently Tupman thinks that to get C from (1) and (2), we need an extra premise, the distributive law, which is what Kriske has begged.

Kriske disagrees. He doesn't think he needs an extra premise to get from (7) and (2) to C. The reasoning by cases helps us *see* that the argument is valid, but it doesn't *add* anything to its validity. If you say otherwise, Kriske thinks, you are begging the question against him.

4 CAN LOGIC BE CHANGED?

In fact, Kriske argues, the very idea of "adopting" a logic is incoherent. Tupman thinks of "logics" as systems of axioms that can be treated as hypotheses to be accepted or rejected on the basis of empirical considerations. However, Kriske maintains that we couldn't possibly *adopt* the logic we already have. The inspiration for his argument comes from Quine's "Truth by Convention"[4] and from Lewis Carroll's famous discussion of Achilles and the Tortoise.[5] One way to put Lewis Carroll's point is that if someone didn't already reason according to *modus ponens*, adding it as an explicit axiom wouldn't help. Suppose someone accepts

A

and also accepts

If *A* then *B*

but for some reason doesn't see that *B* follows. Imagine offering him the following as an explicit principle:

MP: If "*A*" is true and "If *A* then *B*" is true, then "*B*" is true.

Unless the person already grasps *modus ponens*, this won't do any good. He accepts that "*A*" is true and he also accepts that "If *A* then *B*" is true. Let's add, although it's not as trivial as it seems in this context, that he also accepts "'*A*' is true and 'if *A* then *B*' is true." Suppose he also agrees, perhaps accepting your authority, that MP is correct. The problem is that MP is a conditional. To conclude that "*B*" is true, he'll have to reason by *modus ponens*, which is precisely what he wasn't able to do in the first place.

Kriske points out, following Quine, that the same difficulty comes up for universal instantiation. If someone didn't already see, for example, that "All ravens are black" commits her to "Jake (a particular raven) is black," then adding that "All universal statements imply their instances" wouldn't help. The principle itself is a universal statement, and to apply it to a particular case, we would have to infer an instance of it. We run into the same trouble with the rule of conjunction and the principle of non-contradiction; details are left as exercises.

So much for treating logical laws as hypotheses that we might adopt or reject based on the fecundity of their consequences. If by "logic" we mean what we use when we reason, then there's no neutral ground outside logic where we can stand and make judgments about how to draw those consequences.

Kriske briefly discusses two cases in which it might seem that we have allowed or at least considered changes in logic. One is intuitionism. The other is the rejection of the Aristotelian principle that "All P are Q" implies "Some P are Q." In the case of intuitionism, Kriske maintains that the intuitionists didn't reject the rules that applied to the old connectives but rather introduced new connectives. The classical negation of a mathematical statement, in the intuitionists' view, is not guaranteed to be a mathematical statement. Intuitionistic negation can be explained by way of notions we already understand, and it keeps us within mathematics when we reason mathematically. As for Aristotelian logic, modern logicians say that we can get from "All P are Q" to "Some P are Q" *if and only if we assume that there are P's*. But there are cases where "All P are Q" is true even though there aren't any P's. Seeing this isn't "changing logic"; it's recognizing a mistake simply by using ordinary reasoning.

5 KRISKE CONSIDERED

So far, Kriske has argued that

(1) Tupman treats logic as though it were just another theory – just another set of propositions that we accept or reject on the basis of their consequences. However,
(2) that can't be right because it suggests that logic is "up for grabs," when in fact we couldn't consider the consequences of the supposed theory unless we already had logic to do it with – that is, unless we already could reason. Furthermore,
(3) looking at cases like *modus ponens* and universal instantiation makes clear that the very idea of adopting a logic makes no sense. These principles aren't hypotheses; we couldn't adopt them unless we already grasped them. Finally,
(4) there are no good examples of changing logic. In particular, the rejection of Aristotelian logic doesn't count. It's a case of using intuitive reasoning to spot a fallacy.

There are two issues before us. One is whether Kriske is right to think that Tupman's defense of value-definiteness is unsustainable. I think he is, and I will simply assume that from now on. The second issue is whether Kriske has really shown that empirical discoveries couldn't rightly lead us to revise our logical opinions.

5.1 Logic and doxastic practices

If by "logic" we mean something like "correct reasoning," then it would make no sense to think of logic as "just another theory." We need to be able to reason in order to think about anything at all. That said, one suspects that Kriske and the quantum logician may be talking past one another. When Kriske talks about logic, he is talking about a *doxastic practice* in William Alston's sense[6] – a socially established practice of forming and criticizing beliefs. Kriske points out that we don't have a choice about engaging in the practice, and that we have to use logic to justify or criticize logical beliefs. However, Alston reminds us that this is so of other important doxastic practices. We can't avoid using sense experience to form beliefs about the world, but any attempt to justify or criticize either the practice itself or the results of using it will call for relying on things that we learned from the senses – from using the

very practice at issue. In spite of that, specific claims based on sense experience can be treated as hypotheses that could be revised, even though we have to use sense experience to justify the revisions.

This suggests a way to think about challenging logical claims. We need to distinguish between *reasoning* – a doxastic practice – and the theory or discipline in which we attempt to state logical truths and spell out correct forms of inference explicitly. Let's call the output of this discipline "Logic" with a capital "L." Logic in this sense isn't a substitute for the practice of reasoning, but the claims of Logic can be true or false, correct or incorrect and even, perhaps, fecund or barren. Perhaps Tupman could say: we can't put the whole doxastic practice of reasoning up for grabs at once. Nonetheless, we *can* call some of the basic deliverances of reasoning into question, even if we have to reason to do so.

We've already abandoned the hope that we can defend value-definiteness by appeal to Tupman-style quantum logic, but for what comes later it will be helpful to bracket that concession and reconsider the exchange that Kriske imagines himself having with Tupman. Kriske argues, reasoning by cases, that *A* and *B* can't both have values. He imagines Tupman accusing him of begging the question – of omitting a premise (the distributive law) that he needs for his argument to be valid. Kriske replies that Tupman would be begging the question against him; as Kriske sees it, he doesn't need the extra premise. But consider the case of Aristotelian logic. Suppose we insist that that the principle of subalternation is false – that from "All dogs are mammals," it doesn't follow that some dogs are mammals. We insist that the conclusion calls for an extra premise: dogs exist. Imagine the Aristotelian replying that he needs no such premise and hasn't begged any question. Subalternation is valid, he claims, and we modern logicians are begging the question against him if we claim the he needs an extra premise. How would the debate proceed?

The first point is that it actually *could* proceed. The Aristotelian might insist that in cases where "P" is empty, "All P are Q" isn't true. After all, both "All Martians are Americans" and "All Martians are non-Americans" sound odd, even though modern logicians say that both are true. Of course "All Jedi Knights have superhuman powers" seems to be true, but so does "Some Jedi Knights turned to the dark side of the Force." Since the latter hardly entails that there really are Jedi Knights, the Aristotelian could argue that "All Jedi Knights have superhuman powers" was never *literally* true in the first place. If so, it doesn't count against the claim that universal categoricals are false if their subject terms are empty.

The debate could continue. We could point out to the Aristotelian that if "Some P are Q" entails the existence of Ps, as he presumably would agree, then he will have to give up either the principle of conversion for universal negatives or the principle of obversion. (Hint: start with the banal truth "No Canadians are Martians." Then convert, obvert and take the subaltern). Modern logicians have decided that things go more smoothly if we adopt the Boolean interpretation of categorical statements. Nonetheless, no matter what solution we settle on, it will make for some intuitive strain. If there's something to negotiate in the case of the principle of subalternation, Tupman might insist that we can also negotiate in the case of the distributive law.

5.2 Self-presupposing principles

There's a particular difficulty with this reply that we'll get to below. Meanwhile, we come up against the third of Kriske's four points: we seem to be assuming that the distributive law is a hypothesis that can be adopted or rejected based on its consequences. But the discussion of *modus ponens*, universal instantiation and so on was meant to show that the idea of adopting a logical law or logical axiom makes no sense to begin with.

The self-presupposing quality of these principles is striking. However, Tupman could point out that the distributive law doesn't have this peculiarity. It's hard to see that no one could adopt it unless he already grasped it. Furthermore, even if it weren't possible to *adopt* the distributive law, *rejecting* it might still be possible. Though the examples are controversial, it has been argued (most famously by Van McGee[7]) that *modus ponens* doesn't hold in all cases. Whatever one makes of the examples, it's no reply to point out that no one could adopt *modus ponens* if he didn't already grasp it. Likewise, for all the Carroll/Quine/Kriske examples show, the distributive law may be a principle that we could reject. Tupman would say that empirical discoveries have uncovered exceptions to what had looked like a logical truth.

5.3 Internal vs. external

Kriske would insist that we've missed the point. The issue over subalternation is entirely an in-house squabble that never takes us outside the doxastic practice of reasoning. Tupman's case against the distributive law is extramural. He isn't arguing that there's an intuitive objection to the distributive law. He's claiming that if we give it up, we gain a certain *extra-logical* benefit: a realist interpretation of quantum mechanics. That, Kriske would insist, misses the point that logic is all about *reasoning*.

6 QUANTUM *LOGIC?*

Kriske's view of logic is something that we might call *Intuitivism*: claims about logical truth and logical consequence must be grounded in intuitive reasoning. And though just what might count as an intuitive consideration isn't easy to say, appeals to contingent empirical facts don't make the grade.

There's a related point. If Kriske is insisting that by its nature, logic is *a priori* (a matter of "reasoning" and "intuition") then quantum logic seems excluded from the start. Quantum logicians are making claims about physical reality, but they don't claim that the structure of physical reality is something we can know *a priori*.

That's surely right; we can't figure out the structure of the world just by reasoning. Nonetheless, I think it still may be possible to meet Kriske on his own terms. Interestingly enough, his discussion of Aristotelian logic provides a hint. According to Kriske, the Aristotelian's mistake was to overlook something: the possibility that a universal categorical might be true even though its subject term is empty. What the quantum logician must say is that the classical logician has also overlooked some possibility.

6.1 Minimal Quantum Logic

Consider the following three theses:

> *I* The propositions of Quantum Logic (call them Q-propositions) are ascriptions of values to quantum mechanical quantities or logical constructions of such propositions.
> *II* Not every quantum mechanical quantity has a value
> *III* When a quantum mechanical quantity lacks a value, there are true disjunctive Q-propositions whose disjuncts are not true.

I is a stipulation. It says that this is what Quantum Logic, as understood here, will be constructed from. *II* is widely accepted even by people who want nothing to do with Quantum Logic. *III* is the most contentious of the three theses. Though we'll need to say more, we can use an example to provide some motivation. Suppose that a spin-one particle is in the state $|Sz = 0\rangle$. In that case, *S*x doesn't have a value; none of the propositions

$$Sx = +1, Sx = 0, Sx = -1$$

is true – or so it's reasonable to believe. However, there's a case to be made for saying that $(Sx)^2$ does have a value – a value of 1 – even though neither "$Sx = +1$" nor "$Sx = -1$" is true. On the view we're considering, the fact that $(Sx)^2 = +1$ will be the same fact as the one expressed by the disjunction

$$Sx = +1 \vee Sx = -1.$$

If one is true, so is the other.

6.1.1 The distributive law revisited Our theses *I* through *III* don't give us full-blown Quantum Logic, but they're enough to make sense of how someone might think that the distributive law could fail. Suppose that P is a true Q-proposition. Suppose that $Q \vee R$ is also a true Q-proposition, but with disjuncts that aren't true (whether or not we say that they're false is another matter; more on that below.) In this case, the conjunction

$$P \wedge (Q \vee R)$$

will be true, but neither of the propositions

$$P \wedge Q, P \wedge R$$

will be. (We'll leave aside for the moment the question of whether these "propositions" are even well-defined.) The "intuitive" explanation is that the Kriske-style classical logician has overlooked something: the possibility of a true disjunction with disjuncts that aren't true.

It's worth stressing that this account of how the distributive law fails isn't what Tupman, let alone Putnam, had in mind. The value-definiteness thesis is gone. That means that some of Kriske's criticisms of Tupman are no longer relevant. However, Kriske might still insist that

$$(P \wedge Q) \vee (P \wedge R)$$

simply follows from

$$P \wedge (Q \vee R)$$

He might also say that the "possibility" he's accused of overlooking – that a disjunction might be true when neither of its disjuncts is – doesn't deserve to be taken seriously. Given the sketchiness of the defense we've offered for *III*, this wouldn't be unreasonable, though we'll have more to say later. But *I* through *III* are not the central claims of Quantum Logic. What's really at stake lies a little deeper.

6.2 The deeper level

Quantum mechanics represents physical quantities in a striking way. The particular feature of structure that Quantum Logic focuses on is the family of relations of necessary equivalence, necessary exclusion, and entailment that quantum mechanics seems to embody. We can illustrate with a familiar example from Kochen and Specker: a spin-one particle and the components of spin in three orthogonal directions x, y and z. Each of the spin matrices Sx, Sy, Sz has three eigenvalues, -1, 0 and $+1$, corresponding to the three possible results of a measurement of the spin component. The squares of each of these matrices, $(Sx)^2$, $(\mathbf{S}y)^2$ and $(Sz)^2$, each have eigenvalues 0 and 1. The distinctive part of the story begins when we introduce the operator

$$H_S = a(Sx)^2 + b(Sy)^2 + c(Sz)^2$$

whose eigenvalues are $x_0 = b + c$, $y_0 = a + c$, and $z_0 = a + b$. Here the vector $|x_0\rangle$ is also an eigenvector of Sx and of $(\mathbf{S}x)^2$, with eigenvalue 0. Corresponding remarks apply to $|y_0\rangle$ and $|z_0\rangle$. A contextualist would say that

$$H_S = x_0, Sx = 0, (Sx)^2 = 0$$

represent distinct propositions that might differ in truth value. Quantum Logic treats these propositions as necessarily equivalent – as picking out the same possible state of affairs. As for necessary exclusion, the vectors

$$|x_0\rangle, |y_0\rangle, |z_0\rangle$$

are mutually orthogonal. The contextualist would say that in spite of this, it's possible for Hs to take the value x_0 and Sy to take the value 0 at the same time. Once again, Quantum Logic treats these propositions as necessarily excluding one another – as

denoting states of affairs such that if it's true that one obtains, it's false that the other does. Finally, the vector $|z_0\rangle$, for example, is a superposition of $|x_+\rangle$ and $|x_-\rangle$. That means that the subspace corresponding to "$Sz = 0$" lies within the subspace corresponding to "$Sx = +1 \vee Sx = -1$". Quantum Logic take the truth of "$Sz = 0\rangle$" to necessitate the truth of "$Sx = +1 \vee Sx = -1$."

Putting all this together gives us a fourth thesis:

> *IV* Each Q-proposition is associated with a subspace of a Hilbert space. (i) If two propositions are associated with the same subspace, the propositions are necessarily equivalent. (ii) If two propositions are associated with orthogonal subspaces, then the truth of one proposition entails the falsity of the other. (iii) If the subspace associated with a Q-proposition Q lies within the subspace associated with the subspace associated with Q', then the truth of Q entails the truth of Q'.

This is the heart of Quantum Logic. The algebraic structure of the theory suggests a particular network of relations among quantum mechanical properties. According to Quantum Logic, these relations are reflected in logical relations among propositions that ascribe properties to the system. Taken in small handfuls, the relations don't lead to any conflict with classical logic. For example, we could describe a classical structure that exhibits the relations of equivalence and exclusion among Hs, Sx, Sy, Sz, $(Sx)^2$, $(Sy)^2$ and $(Sz)^2$. However, as the network grows, we reach a point where a classical structure can't make room for the relations. If we held onto the view that every quantity always has one of its values, this tipping point would be a collapse into incoherence. Quantum Logic tells another story.

7 GLEASON'S THEOREM

Consider a finite algebra B of propositions that obey classical logic. The algebra will be Boolean, and it will contain atoms – maximally informative non-contradictory elements. If we assign the value 1 (i.e., true) to an atom, then the truth value of every other proposition in the algebra is determined. Furthermore, these truth values amount to a measure on the algebra B, with values in the interval [0,1].

Talking about the whole interval [0,1] is a bit coy, but the reason is probably obvious. Suppose that A is an algebra of Q-propositions associated with a finite-dimensional Hilbert space of dimension 3 or greater. Then A also has atoms, and if we assign the truth value 1 to one of these atoms, there is a unique measure on A that assigns each proposition a value in the interval [0,1].[8] This is a consequence of Gleason's theorem[9], and we'll call such values the *Gleason measures* of the propositions. The difference, however, is that in the classical case, all the values are in the set {0,1}; in the quantum case, they fill up the whole interval [0,1].

Let H be a Hilbert space of dimension 3 or greater, and let A be the associated algebra of propositions. (Whether A is a lattice or a partial Boolean algebra is something we don't need to decide at this point.) Suppose that R is a ray in H, and that R is the associated proposition. Now suppose that R is true and let S be a sphere

containing R. By *IV* (iii), the proposition S associated with S is true, but S has many representations. In particular, there are infinitely many disjunctions $S_1 \vee S_2 \vee S_3$, where the S_i correspond to orthogonal rays, and where $S_1 \vee S_2 \vee S_3$ is equivalent to S. By *IV* (ii), if one of the disjuncts is true, then the others are false. Could each such disjunction be true by virtue of the truth of one of its disjuncts?

It's an immediate consequence of Gleason's theorem – or of Kochen and Specker's – that the answer is no. But since R implies each of these disjunctions, this tells us that if Quantum Logic is correct, there must be true disjunctions without true disjuncts.

This need not mean that all the disjuncts are false. Consider

V There are Q-propositions that are neither true nor false.

It seems wrong to say that the propositions corresponding to the components of a superposition are true. However, interference effects are real; the possibilities that correspond to the components of a superposition seem to have an influence on the actual that would be strange if these propositions were simply false. The thought behind *V* is not that we need to make room for vagueness or linguistic indecision, but for the strange way in which the components of a quantum superposition bear on the world.

In any case, if we say that some propositions are neither true nor false, we avoid an unpleasant consequence: true disjunctions all of whose disjuncts are false. But if not true and also not false, then what? Perhaps we don't need a firm answer, but here is one possibility. We could take the Gleason measures induced by the truth of an atom to be *truth-values*. When a proposition's Gleason measure is close to 1, its contribution to the superposition all but swamps the contributions of the other components; when its Gleason measure is close to 0, it makes all but no contribution – and so on. And of course, if the Gleason measures are truth-values, then we can say more about the truth of disjunction. In classical logic, the truth value of an exclusive disjunction is the sum of the truth values of its components. The same would be true for Gleason-measure truth values. In particular, if $P \vee Q$ is a true quantum disjunction with mutually exclusive disjuncts, then the "Gleason truth-values" of the disjuncts will sum to 1. Also, the more complicated rules that apply to non-exclusive disjunctions will be borne out as well, provided the components of the disjunction all belong to a common Boolean algebra. (This, by the way, seems like a reason for preferring partial Boolean algebras to lattices.)

We will remain agnostic about whether Gleason measures are truth values; the issues would take us too far afield. Nonetheless, Gleason measures do encode real features of the system. We'll say more about this later.

8 ANSWERING KRISKE

We have the materials for answering Kriske on his own terms. The claim is that if we rest with classical logic, we've overlooked something. We described it as the possibility that disjunctions might be true even though their disjuncts aren't. The more basic notion is incompatibility or, as I prefer, incommensurability. Two propositions are incommensurable if they don't belong to a common Boolean algebra. This general

notion applies in principle to a broader class of situations than the purely quantum mechanical. If we concentrate on quantum mechanics, and if we agree that some Q-propositions are neither true nor false, then incommensurability amounts to this: two Q-propositions are incommensurable if (a) neither implies the other, and (b) the truth of one rules out the falsity of the other. Notice that (b) isn't possible classically unless (a) is false; in classical logic, if a proposition X rules out the falsity of a proposition Y, then X implies Y.

For Q-propositions, (a) and (b) can be restated in terms of Gleason measures: P and Q are incommensurable if (a) there are Gleason measures that assign 1 to P but not to Q and vice-versa, and (b) every Gleason measure that assigns 1 to P assigns Q a value strictly greater than 0 and vice-versa.

The claim, then, is that the classical logician has overlooked the possibility that propositions can be incommensurable. However, Kriske's model of a case in which a logician has overlooked something is the rejection of Aristotelian logic, where what was overlooked could be uncovered simply by reasoning. Is this notion of incommensurability likewise something that we could have come up with simply by reasoning?

Perhaps. We can imagine a story like the one in Paper Four of *Interpreting the Quantum World*.[10] There, Jeff imagines a bright student who invents quantum mechanics as a thought experiment while thinking about Hilbert space. We could tell a similar tale for the interpretation we're offering here. On such a story, the empirical discovery relevant to Quantum Logic would not be the discovery that incommensurability is a coherent notion, but rather the discovery that there actually *are* empirically significant incommensurable propositions. This is something that couldn't have been known *a priori*, and so it couldn't have been known *a priori* that Classical Logic is inadequate for describing physical reality. But just as it could be and arguably was known *a priori* that geometry didn't *have* to be Euclidean, so it could have been, though wasn't, known *a priori* that *I* through *V* could all be true. Or so the quantum logician would say.

In actual fact, the idea that Classical Logic might be inadequate wasn't dreamt up as an exercise in abstract mathematics. It was a response to empirical discoveries rather than a speculation that guided them. However, the relevant question for answering Kriske is whether the quantum logician's proposal amounts to a coherent thesis. This is a conceptual question, even though it almost certainly would never have arisen but for certain scientific developments. Given Tupman's understanding of Quantum Logic, Kriske was right to accuse him of incoherence. However, Kriske's specific criticisms were directed at the claim that whenever a disjunctive Q-proposition is true, one of its disjuncts is true. Those criticisms have no force here. His more general line of attack was that the idea of "adopting a logic" is incoherent and that logic is ultimately a matter of reasoning or "intuition." The reply is that there is an issue for "reasoning" or "intuition," or perhaps better, philosophical reflection here: whether the quantum logical proposal is coherent. Perhaps it's not. But it won't do, for instance, to insist that from $P \wedge (Q \vee R)$ it follows that one of $(P \wedge Q)$, $(P \wedge R)$ is true. The quantum logician claims that this overlooks a real possibility: the possibility that the pairs of propositions (P, Q) and (P, R) might be incommensurable.

8.1 Disjunction defined?

Some may think (Tim Maudlin, for example[11]) that disjunction is *defined* by the requirement that for a disjunction to be true, at least one of its disjuncts must be true. However, this definition operates against the background assumption that all propositions have truth values. When that assumption falls away, it's no longer so clear that this is the best understanding of disjunction. Notice that there will be a "truth-maker" for a quantum disjunction. It will be the truth of the proposition that implies it. Also, there will be a true disjunction that *is* true by virtue of the truth of one of its disjuncts, and that is equivalent to the anomalous disjunction with its non-true disjuncts. Furthermore, if either disjunct of our anomalous disjunction *were* true, then that disjunct would be a truth-maker for this disjunction. In other words, the quantum disjunction can be made true by one of its disjuncts; in the right circumstances, it behaves like a classical disjunction. What the quantum logician adds is that there are also circumstances not hitherto dreamt of in our philosophies.

9 FROM LOGIC TO THE LAB

What's been said so far about the coherence of Quantum Logic is at best a sketch of a defense. However, suppose the sketch could be filled in. We come now to a harder question. Suppose we allow that the logical relationships among properties *could* be as the quantum logician says. How could we know – or at least reasonably believe – that there really are systems with that sort of structure?

The answer to that question surely turns at least in part on another: what would we expect to see if we encountered a system whose property structure fit Quantum Logic? Even if it's possible for propositions to be incommensurable, it's not necessary that any actually are; the non-classical features of Quantum Logic *could* fail to fit the real world. To have reason to believe that the world has quantum-logical features, we would have to have reason to think that those features would make a detectable difference to the way things behave. And so we need to ask: what would that difference be?

At this level of generality, there is no clear answer. To find out anything about a system, we have to interact with it, and unless we know something about the sorts of interactions that can take place, we have no basis for any expectations. What would we need to assume about systems supposedly described by Quantum Logic for empirical questions about them to have any content? I will take it for granted that we assume each Boolean subalgebra to correspond to an observable. We would also need to assume that we can prepare systems in such a way that certain Q-propositions are true of them – that we can prepare states, in effect. And to have any assurance of that, we would also have to assume that *if* a system is prepared with a certain property, there are reliable ways of making it display the property. All of this needs more spelling out. For the sake of brevity, I will gesture to the assumptions that Simon Saunders makes use of in the first three sections of his "Derivation of the Born Rule from Operational Assumptions."[12] However, a tricky issue remains.

9.1 Two kinds of contextualism

A Q-proposition will typically belong to many Boolean subalgebras. Quantum logic is non-contextual in that it counts the proposition as picking out the same state of affairs regardless of which Boolean subalgebra we associate it with. This is, as it were, *ontological non-contextualism*. However, there is a further *empirical* issue about contextualism.

Suppose Q is a quantity with distinct values $q_1, q_2 \ldots q_n$ and that R is another quantity with values $r_1, r_2, \ldots r_n$. Suppose that none of the propositions "$Q = q_i$", "$R = r_i$" are true and none are false. (On the standard picture, this would amount to supposing that the state vector is given by

$$|\psi\rangle = \Sigma_i c_i |q_i\rangle = \Sigma_i d_i |r_i\rangle$$

where none of the coefficients are zero.) Finally, suppose that $|q_1\rangle\langle q_1| = |r_1\rangle\langle r_1|$.

First, consider a measurement of Q. What should we expect?

We know what to expect *in fact*: repeated measurements of Q in state $|\psi\rangle$ should yield distributions of results that accord with the Born rule. However, the question is what we should expect if we look at things with an eye to figuring out what a quantum logical world would be like.

Suppose we can expect to get some result or other – a macroscopic event that betokens one of the eigenvalues q_i. And suppose we want to assign probabilities to such results. How should we do it?

It won't do simply to appeal to the Born Rule. Our assumption is that the various Q-propositions are related as Quantum Logic says they are. But Quantum Logic is an account of relations of equivalence, exclusion and implication. That doesn't immediately tell us anything about experimental probability.

It might seem that we can easily bridge the gap. As we have already pointed out, the quantum logical structure , together with the assumption that the proposition associated with $|\psi\rangle\langle\psi|$ is true (call it P), yields a Gleason measure on all the propositions. This measure is unique; there is no other way to assign numbers in [0,1] simultaneously to all the Q-propositions in such a way that the numbers yield measures on each of the Boolean subalgebras. Moreover, the Gleason measure of a proposition will be the very number given by the Born Rule. In this case, the Gleason measure of proposition $Q = q_1$ will be $|\langle\psi|q_1\rangle|^2$, and since $|q_1\rangle = |r_1\rangle$, this will also be the Gleason measure of $R = r_1$. Can't the quantum logician simply treat this as a probability?

Not without further assumptions. Let's grant that when P is true, the system has some feature represented by the fact that $Q = q_1$ has Gleason measure $|\langle\psi|q_1\rangle|^2$ – in this case, the same feature as the one represented by the fact that $R = r_1$ has the Gleason measure $|\langle\psi|r_1\rangle|^2$. The question, however, is what understanding of this feature Quantum Logic is entitled to. It's not hard to see how it might *fail* to be a probability.

A measurement of Q might elicit the property associated with $Q = q_1$. Furthermore, since this is the same property as the one associated with $R = r_1$, that would also

count as eliciting the property associated with $R = r_1$. Parallel comments apply to a measurement of R. But even though Quantum Logic is ontologically non-contextual, two different probabilistic ways of eliciting one and the same property *could* yield two different probabilities. Put another way, even though Quantum Logic treats quantum quantities as ontologically non-contextual, it doesn't rule out the possibility that they are *empirically* contextual.[14] Given that we've rejected the value-definiteness thesis, Quantum Logic will have to say that a measurement typically doesn't just reveal something; it induces a change in the system. There's nothing incoherent in the thought that the way in which the change is induced might affect the probabilities for one and the same micro-event to occur.

9.2 Trimming the context tree

In order to know what to expect if Quantum Logic is correct, we need to assume more than that quantum mechanical propositions are related as Quantum Logic says they are. One obvious additional assumption is that for ideal measurements, the only properties that bear on the empirical probabilities are the ones encoded in the Quantum Logical algebra of propositions – that those are what ideal measuring instruments respond to. This is hardly an *ad hoc* move. Making such an assumption amounts to assuming that the relations embodied in the Quantum Logical algebra of propositions are fundamental for determining how quantum systems will behave; there are no further "hidden variables." This fits with the idea that quantum theory is a principle theory whose fundamental constraints are given by the Quantum Logical algebra of propositions – an idea that has long been part of Quantum Logic.[14] If we make this assumption, then Gleason's theorem guarantees that the only possible assignments of probabilities are the ones that accord with the Born rule. Though more needs to be said, Quantum Logic at least offers the promise of a coherent, attractive foundation for thinking about quantum probability. The probabilities emerge from the most basic features of the quantum quantities: their logical relationships to one another.

10 THE QUANTUM LOGICIAN'S CONUNDRUM

We've described a version of Quantum Logic that avoids the incoherence of Tupman's approach but still counts as realist: it sees Quantum Logic as an hypothesis about the way the world is structured. It also takes seriously Kriske's challenge that logic must be grounded in *reasoning*. It meets the challenge by claiming that classical logic has overlooked a coherent possibility: a disjunction could be true even though none of its disjuncts are. As a mere abstract claim, this would have little to recommend it. However, we pursued the idea that at its heart, Quantum Logic is a view about the way in which quantum-mechanical properties are related to one another. The thesis about disjunction is grounded in this deeper picture. Furthermore, Quantum Logic offers the beginnings of an appealing treatment of probabilities.

All of this *seems* to add up to an answer to the question "How would things behave if Quantum Logic were correct?" The answer seems to be: they would behave the

way that quantum theory, as usually understood, says they would. In fact, this answer is problematic.

10.1 How Quantum Mechanical is a Quantum Logical world?

What has been said so far leaves some large questions. For one thing, nothing has been said about dynamics. Measurement aside, dynamical transformations are usually thought of in Quantum Logic as automorphisms on the algebra of propositions; every unitary transformation on a Hilbert space induces such an automorphism. However, we can't leave measurement aside, and Quantum Logic as presented here can't avoid the measurement problem. Measurements are stochastic changes in the properties of the systems, and they can't be modeled by automorphisms on the algebra. Quantum Logic has nothing to say about what induces those changes. Worse still, if the explanation for non-unitary change is some additional variable, it will no longer be clear that empirical probabilities should depend only on which Q-propositions are true before the measurement and on what's encoded in the algebra of propositions. This puts Quantum Logic's account of quantum probability at risk.

Of course, it's not clear that quantum mechanics itself has much to say about what why measurements have results. The measurement problem, after all, is the problem of explaining how *quantum mechanics* can provide a satisfactory account of measurement. Perhaps the Quantum Logician can punt on this issue. It's not clear that in order to be viable, Quantum Logic has to answer all interpretive questions. All the quantum logician need claim is that the structures Quantum Logic posits are *part* of the story of why quantum systems behave as they do.

There's an obvious related issue. Since Quantum Logic posits indefinite values, the problem of Schrödinger's cat looms on the horizon. Once again, the difficulty isn't peculiar to Quantum Logic, but the rejection of value-definiteness means that Quantum Logic can't dodge the problem in any easy or obvious way. Still, we might say, although Quantum Logic must be consistent with some acceptable solution to these problems, it needn't contain the solution itself.

And then there's locality. If a pair of electrons is in the singlet state, then the quantum logician is committed to saying that none of the local spin quantities have values. However, after a spin measurement on one of the systems, what was once indefinite on the distant system will become definite. Something has changed "there" because of something that happened "here." Quantum Logic may be able to avoid ontological non-locality, but it's hard to see how it can steer clear of non-local causal influences. Furthermore, we have the familiar problem of selecting the hyperplane on which the change occurs.

Once again, we have a difficulty that's hardly unique to Quantum Logic. But that excuse may be wearing thin. Quantum Logic appears to have nothing to contribute to the problem of explaining why measurements have results; it practically ensures that we will face the problem of Schrödinger's cat; and it seems to be on a collision course with special relativity. But measurements *do* have results, superposed cats appear to be mythical beasts, and conflict with special relativity is to be avoided where possible.

Although we pointed out that there are lingering issues about contextualism, the area where Quantum Logic seems to show the most promise is in understanding quantum probability. However, it's not clear that Quantum Logic has any real advantage here. Recent work by Deutsch[15] and Wallace[16] on probability in the Everett interpretation has been extended by Simon Saunders[17] to all interpretations that treat different ways of performing measurements as equivalent whenever they are unitarily equivalent. Saunders shows, generalizing Deutsch's result, that with this assumption, we can derive the Born rule from what he refers to as operational assumptions. The proof is compact and elegant; no need for Gleason's theorem.

We've arrived at the conundrum. For it to be plausible that Quantum Logic is a coherent thesis, there has to be a good answer to the question of what the world would be like if it were quantum logical. We know that the world acts the way that *quantum mechanics* says it does, and we know that Quantum Logic fits neatly into the standard mathematical apparatus that quantum mechanics uses. But quantum mechanics is not just its mathematics; it's also a set of techniques and practices for applying the math. We know that trying to think of that mathematical apparatus as a depiction of the world leads to the frustratingly hard problems of interpretation that have kept workers in the foundations of physics employed for decades. It may be that the usual mathematical apparatus is nothing but the guts of a highly successful prediction machine that *can't* be taken at face value. And it may be that Quantum Logic is the purest expression of what makes the standard theory so hard to interpret!

11 INCONCLUSIVE CONCLUDING THOUGHTS

The version of Quantum Logic under consideration here is an attempt to follow the realist instinct that motivated Western Ontario-style quantum logic thirty odd years ago. What's gone are the twin commitments to definite values and to the thesis that measurements simply reveal. The realism that remains consists in two things: first, the claim that quantum properties really embody the logical relations that Quantum Logic says they do, and second the claim that this fact helps explain why quantum systems behave as they do. And while bivalence is gone, this version of Quantum Logic takes the Gleason measures of propositions to be real features of the system. If we were to take Gleason measures as truth values, then even though bivalence itself would be gone, we would have a definite though non-standard realist understanding of truth.

This is by no means the only approach to quantum logic (lower-case to indicate the generic.) Much work done under the heading "quantum logic" is frankly operational and makes no radical claims about the structure of properties.[18] More recently, William Demopoulos has offered an understanding of quantum logic according to which the *logical* relations among quantum propositions aren't represented by the structure referred to here as Quantum Logic, but by a much less constraining structure that allows every quantum mechanical proposition to be determinately true or false. However, on Demopoulos's view complete knowledge of a quantum system is impossible in principle; the structure that we have been calling Quantum Logic has

epistemic rather than alethic significance. It represents constraints on our *knowledge* of the quantum world.[19]

Quantum mechanics is strange business and whatever the true story of the quantum world may be, it's safe to say it's weird. Wildly different interpretations abound; none can claim wide allegiance. Worse still, it's far from clear how we should even go about deciding among the competitors. Bohmian mechanics is consistent, far as I know. Is it true? How would we decide? The Everett interpretation is probably consistent. It *may* even be able to make sense of the probabilities, though I have my doubts.[20] But even if such doubts can be resolved, many of us find it hard to imagine actually believing that the picture is correct. However – and not helpfully for getting at the truth – all of this may be a matter of taste. Quantum Logic invokes its own incredulous and irrefutable stares.[21]

So there we are. Perhaps for purely sentimental reasons, I'd like to think that Quantum Logic, understood in a realist way, is a coherent conjecture, and that it can make some genuine explanatory contribution to our understanding of quantum systems. I'd like to think this; I'm not quite ready to say it isn't so. But if the True Believers were asked to stand, I fear I'd be huddled in the corner with those of flickering faith.

NOTES

[1] Kochen, S. and Specker, E. P., "The problem of hidden variables in Quantum Mechanics," *Journal of Mathematics and Mechanics* **17** (1967) 59–67.

[2] Our way of writing the disjunction is a bit misleading. The statement $Q = \mathrm{q_i}$ will correspond to the same subspace as various other statements, e.g., $R = \mathrm{r_j}$, and for the contradiction to emerge, each of these statements must be taken to represent the same proposition and expressed accordingly.

[3] Putnam, Hilary, "Is logic empirical?" in R. Cohen and M. P. Wartofski (eds.), *Boston Studies in the Philosophy of Science* **5** (Dordrecht, Holland: D. Reidel, 1968).

[4] Quine, Willard van Orman *The Ways of Paradox and Other Essays* (Revised and Expanded Edition), Cambridge: Harvard University Press, 1976. 102, ff.

[5] Carroll, Lewis, "What the Tortoise Said to Achilles," *Mind* **4** (1985) 278–280.

[6] Alston, William, *Perceiving God*, Ithaca NY: Cornell University Press, 1991.

[7] McGee, Van, "A Counterexample to Modus Ponens." *Journal of Philosophy* (September 1985), 82: 462–471.

[8] By calling this a measure, we mean that it is defined on subspaces of the Hilbert space, that the measure of the whole space is 1, and that if S and S' are orthogonal subspaces, then the measure of their span $S \vee S'$ is the sum of the measures of S and S'.

[9] Gleason, Andrew M. "Measures on the closed subspaces of a Hilbert space," *Journal of Mathematics and Mechanics*, **6** (1957) 885–893.

[10] Bub, Jeffrey, *Interpreting the Quantum World*, Cambridge: Cambridge University Press, 1997.

[11] Maudlin, Timothy, "The Tale of Quantum Logic" (forthcoming).

[12] Saunders, Simon, "Derivation of the Born Rule from Operational Assumptions," *Proceedings of the Royal Society, London* **A 460** (2004) 1–18.

[13] This distinction is rather like the distinction that Heywood and Redhead make between ontological and environmental contextualism. One notable difference is that Heywood and Redhead's distinction is set in the context of a discussion of local hidden variables. See Heywood, P. and Redhead, M. L. G., "Non-locality and the Kochen-Specker paradox," *Foundations of Physics* **13** (1983) pp. 481–499. See also Redhead, Michael, *Incompleteness, Nonlocality and Realism*. Oxford: Clarendon Press (1987) ch. 6.

[14] See, for example, Bub, Jeffrey, *The Interpretation of Quantum Mechanics*. Dodrecht, Hollad; Reidel (1974) p. viii, 143.

[15] Deutsch, D. "Quantum Theory of probability and decisions." *Proceedings of the Royal Society of London* **A** 455 (1999) 3129–3137.

[16] Wallace, David, "Everettian rationality." *Studies in History and Philosophy of Modern Physics* 34 (2003) 87–105.

[17] Loc. cit.

[18] For a review of various approaches to quantum logic, see Wilce, Alexander, "Quantum Logic and Probability Theory, *Stanford Encyclopedia of Philosophy*, http://plato.stanford.edu/entries/qt-quantlog/#2.

[19] Demopoulos, William, "Elementary propositions and essentially incomplete knowledge: a framework for the interpretation of quantum mechanics," *Noûs* **38**: 1 (2004) 86–109. The logical relations, on Demopoulos's view, are represented by the free partial Boolean algebra whose structure is isomorphic to the partial Boolean algebra of subspaces of two-dimensional Hilbert space.

[20] I don't question the mathematics of David Wallace's improvements on Deutsch's argument. (See the pieces by Deutsch and Wallace cited above.) I do, however, harbor grave doubts about whether a rational agent is in any way constrained to treat possibilities that have very different consequences for how the branching will unfold as though they were equivalent for purposes of probability.

[21] With apologies to David Lewis, who famously remarked, when confronted with certain critics of his modal realism, that it's hard to refute an incredulous stare. For Lewis's methodological discussion of the incredulous stare, see Lewis, David K., *On the Plurality of Worlds*. New York: Blackwell (1986) p. 133 ff.

BIBLIOGRAPHY OF THE PUBLICATIONS OF JEFFREY BUB TO 2006

1966

'A Proposed Solution of the Measurement Problem in Quantum Mechanics by a Hidden Variable Theory,' *Reviews of Modern Physics* **38**, 453–469 (1966). (With David Bohm.)

'A Refutation of the Proof by Jauch and Piron that Hidden Variables can be Excluded in Quantum Mechanics, *Reviews of Modern Physics* **38**, 470–475 (1966). (With David Bohm.)

1968

'On Hidden Variables – A Reply to Comments by Jauch and Piron and by Gudder,' *Reviews of Modern Physics* **40**, 235–236 (1968). (With David Bohm.)

'Miller's Paradox of Information,' *British Journal for the Philosophy of Science* **19**, 63–67 (1968). (With M. Radner.)

'Hidden Variables and the Copenhagen Interpretation – A Reconciliation,' *British Journal for the Philosophy of Science* **19**, 185–210 (1968).

'The Daneri-Loinger-Prosperi Quantum Theory of Measurement,' *Il Nuovo Cimento* **57B**, 503–520 (1968).

Review of *Quantum Theory and Reality*, M. Bunge (ed.); *Philosophy of Science* **35**, 425–429 (1968).

1969

'What is a Hidden Variable Theory of Quantum Phenomena?,' *International Journal of Theoretical Physics* **2**, 101–123 (1969).

1970

'Comment on the Daneri-Loinger-Prosperi Quantum Theory of Measurement,' in *Quantum Theory and Beyond*, T. Bastin (ed.) (Cambridge: Cambridge University Press, 1970), pp. 65–70.

Review of *The Philosophy of Quantum Mechanics*, D.I. Blokhintsev; *Philosophy of Science* **37**, 156–158 (1970).

Review of *Quantum Theory and the Philosophical Tradition*, A Petersen; *Philosophy of Science* **37**, 156–158 (1970).

1973

'Under the Spell of Bohr,' *British Journal for the Philosophy of Science* **24**, 78–90 (1973).

'On the Possibility of a Phase-Space Reconstruction of the Quantum Statistics: A Refutation of the Bell-Wigner Locality Argument,' *Foundations of Physics* **3**, 29–44 (1973).

1974

The Interpretation of Quantum Mechanics (Dordrecht: Reidel, 1974).

'Reply to Professor Causey,' in *The Structure of Scientific Theories*, F. Suppe (ed.) (Evanston: University of Illinois Press, 1974), pp. 402–408.

W. Demopoulos and I. Pitowsky (eds.), Physical Theory and its Interpretation, 273–280.

'On the Completeness of Quantum Mechanics,' in *Contemporary Research in the Foundations and Philosophy of Quantum Theory*, C.A. Hooker (ed.) (Dordrecht: Reidel, 1974), pp. 1–65.

'The Interpretation of Quantum Mechanics,' in *Boston Studies in the Philosophy of Science* Vol. XIII, R.S. Cohen and M. Wartofsky (eds.) (Dordrecht: Reidel, 1974), pp. 92–122. (With W. Demopoulos.)

Review of *The Conceptual Foundations of Contemporary Relativity Theory*, J.C. Graves; *Philosophy of Science* **41**, 431–433 (1974).

1975

'Popper's Propensity Interpretation of Probability and Quantum Mechanics,' in *Minnesota Studies in the Philosophy of Science* Vol.VI, G. Maxwell and R.M. Andersen (eds.) (Minneapolis: University of Minnesota Press, 1975), pp. 416–429.

Review of *Foundations of Physics*, M. Bunge; *Philosophia* **5**, 352–356 (1975).

1976

'Randomness and Locality in Quantum Mechanics,' in *Logic and Probability in Quantum Mechanics*, P. Suppes (ed.) (Dordrecht: Reidel, 1976), pp. 397–420.

'The Statistics of Non-Boolean Event Structures,' in *Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science*, W.L. Harper and C.A. Hooker (eds.) (Dordrecht: Reidel, 1976), pp. 1–16.

'Hidden Variables and Locality,' *Foundations of Physics* **6**, 511–526 (1976).

Critical Study: *Paradigms and Paradoxes*, R.G. Colodny (ed.), in *Philosophia* **6**, 511–526 (1976). (With W. Demopoulos. Review article.)

1977

'Von Neumann's Projection Postulate as a Probability Conditionalization Rule in Quantum Mechanics,' *Journal of Philosophical Logic* **6**, 381–390 (1977).

'What is Philosophically Interesting About Quantum Mechanics?' in *Proceedings of the Fifth International Congress on Logic, Methodology, and Philosophy of Science: Part Two: Foundational Problems in the Special Sciences*, R. Butts and J. Hintikka (eds.) (Dordrecht: Reidel, 1977), pp. 69–79.

1978

'Conditional Probabilities in Quantum Mechanics,' in *The Logico-Algebraic Approach to Quantum Mechanics* Vol. II, C.A. Hooker (ed.) (Dordrecht: Reidel, 1978), pp. 209–226.

'Non-Local Hidden Variable Theories and Bell's Inequality,' in *PSA 1978*, Vol. 1, 45–53 (1978), P.D.Asquith and I. Hacking (eds.) (East Lansing, Michigan: Philosophy of Science Association, 1978).

1979

'Some Reflections on Quantum Logic and Schrodinger's Cat,' *British Journal for the Philosophy of Science* **30**, 27–39 (1979).

'The Measurement Problem of Quantum Mechanics,' in *Problems in the Foundations of Physics*, G. Toraldo di Francia (ed.) (Dordrecht: Reidel, 1979), pp. 71–124. (Proceedings of the Enrico Fermi International School of Physics, Varenna, 1977, LXXII Course.)

1980

'Comment on W. Demopoulos: "Locality and the Algebraic Structure of Quantum Mechanics,"' in *Studies in the Foundations of Quantum Mechanics*, P. Suppes (ed.) (East Lansing, Michigan: Philosophy of Science Association, 1980), pp. 149–153.

Review of *Quantum Logic*, P. Mittelstaedt; *Philosophy of Science* **47**, 332–335 (1980).

1981

'What Does Quantum Logic Explain?' in *Current Issues in Quantum Logic*, E. Beltrametti (ed.), 'Ettore Majorana' International Science Series: Physical Sciences, Vol 8 (New York: Plenum Press, 1981), pp. 89–100.

'Hidden Variables and Quantum Logic – A Skeptical Review,' *Erkenntnis* **16**, 275–293 (1981).

'Complementarity,' in *Encyclopedia of Physics*, R.G. Lerner and G.L. Trigg (eds.) (New York: Addison-Wesley, 1981), pp. 138–139.

1982

'Quantum Logic, Conditional Probability, and Interference,' *Philosophy of Science* **49**, 402–421 (1982).

Review of *Readings from the New Book on Nature: Physics and Metaphysics in the Modern Novel*, R. Nadeau; *Philosophy of Science* **49**, 480–481 (1982).

1983

Review of *Physics and Philosophy: Selected Essays*, H. Margenau; *Philosophy of Science* **50**, 515–516 (1983).

1985

'On the Nature of Randomness in Quantum Mechanics, or How to Count Quantum Logically,' in *Recent Developments in Quantum Logic*, P. Mittelstaedt and E.W. Stachow (eds.) (Mannheim: Bibliographisches Institut, 1985), pp. 45–59.

'On the Non-Locality of Pre- and Post-Selected Quantum Ensembles,' in *Symposium on the Foundations of Modern Physics: 50 Years of the Einstein-Podolsky-Rosen Gedankenexperiment*, P. Lahti and P. Mittelstaedt (eds.) (Singapore: World Scientific Publishing Co., 1985), pp. 333–341.

Critical Notice of Sir Karl Popper's *Postscript to The Logic of Scientific Discovery*, *Canadian Journal of Philosophy* **15**, 539–552 (1985). (With I. Pitowsky.)

Review of *How the Laws of Physics Lie*, N. Cartwright; *Canadian Philosophical Reviews* V, 104–107 (1985).

1986

'Curious Properties of Quantum Ensembles which have been both Pre- and Post-Selected,' *Physical Review Letters* **56**, 2337–2340 (1986). (Senior author with H. Brown.)

1988

'How To Kill Schrodinger's Cat,' in *The World View of Modern Physics: Does it Need a New Metaphysics?* R. Kitchener (ed.) (Albany: SUNY Press, 1988), 59–74.

'From Micro to Macro: A Solution to the Measurement Problem of Quantum Mechanics,' *PSA 1988*, A. Fine and J. Leplin (eds.) (East Lansing, Michigan: Philosophy of Science Association, 1988), pp. 134–144.

'On the Methodology of Single-Case Studies in Cognitive Neuropsychology,' *Cognitive Neuropsychology* **5**, 565–582 (1988). (With D. Bub.)

'How to Solve the Measurement Problem of Quantum Mechanics,' *Foundations of Physic*s **18**, 701–722 (1988).

Review of *Images of Science*, P.M. Churchland and C.A. Hooker (eds.); *Foundations of Physics Letters* **1**, 395–399 (1988). (With D. MacCallum.)

1989

'On Bohr's Response to EPR: A Quantum Logical Analysis,' *Foundations of Physics* **19**, 793–805 (1989).

'The Philosophy of Quantum Mechanics,' review article of Michael Redhead, *Incompleteness, Nonlocality, and Realism* (Oxford: Clarendon Press, 1988); Peter Gibbins, *Particles and Paradoxes* (Cambridge: Cambridge University Press, 1988); Henry Krips, *The Metaphysics of Quantum Theory* (Oxford: Clarendon Press, Oxford), *British Journal for Philosophy of Science* **40**, 191–211 (1989).

'On the Measurement Problem of Quantum Mechanics,' in M. Kafatos (ed.), *Bell's Theorem, Quantum Theory, and Conceptions of the Universe* (Boston: Kluwer Academic Press, 1989), pp. 7–16.

1990

'On Bohr's Response to EPR: II,' *Foundations of Physics* **20**, 929–941 (1990).

'Philosophia na Qvantovata Mechanica,' *Philosophia Mysal* **4**, 74–93 (1990). (Bulgarian translation of 'The Philosophy of Quantum Mechanics,' 1989.)

Review of *Niels Bohr's Philosophy of Physics*, Dugald Murdoch; *Philosophy of Science* **57**, 344–347 (1990).

1991

'Measurement and "Beables" in Quantum Mechanics,' *Foundations of Physics* **21**, 25–42 (1991).

'The Problem of Properties in Quantum Mechanics,' *Topoi* **10**, 27–34 (1991).

'On States and Probabilities in Quantum Mechanics,' *Proceedings of the Joint Concordia-Sherbrooke Seminar Series on Functional Integration Methods in Stochastic Quantum Mechanics*, *Supplemento ai Rendiconti di Circolo Matematico di Palermo*, Serie II, No. 25 (1991), pp. 109–132.

'Complementarity,' in *Encyclopedia of Physics*, Second Edition, R.G. Lerner and G.L. Trigg (eds.), VCH Publishers, New York, 1991. (N.B. This is a completely revised and refereed version of item 7.)

Review of *Incompleteness, Nonlocality, and Realism: A Prolegomenon to the Interpretation of Quantum Mechanics*, Michael Redhead; *International Studies in Philosophy*, Vol. XXIII, 140–141 (1991).

Review of *The Structure and Interpretation of Quantum Mechanics*, R.I.G. Hughes; *Isis*, 174–175 (1991).

Review of *The Philosophy of Quantum Mechanics: An Interactive Interpretation*, Richard Healey; *Isis*, 606–607 (1991).

1992

'Quantum Mechanics as a Theory of "Beables,"' in A. van der Merwe, F. Selleri, and G. Tarozzi (eds.), *Bell's Theorem and the Foundations of Modern Physics* (Singapore: World Scientific, 1992), pp. 117–124.

'A Quantum Logical Solution to the Measurement Problem of Quantum Mechanics,' *International Journal of Theoretical Physics* **31**, 1857–1871, 1992.

'Quantum Mechanics Without the Projection Postulate,' *Foundations of Physics* **22**, 737–754, 1992.

'EPR,' *Foundations of Physics* **22**, 313–332, 1992. (With A. Hajek.)

1993

'Measurement: It Ain't Over Till It's Over,' *Foundations of Physics Letters* **6**, 21–35, 1993.

'Measurement and Objectivity in Quantum Mechanics,' in H.D. Doebner, W. Scherer, F. Schroeck, Jr. (eds.), *Classical and Quantum Sysytems – Foundations and Symmetries*, *Proceedings of the 2nd International Wigner Symposium*, Goslar, Germany (Singapore: World Scientific, 1993), p. 9–18.

'Non-Ideal Measurements,' in P. Busch, P. Lahti, and P. Mittelstaedt (eds.), *Symposium on the Foundations of Modern Physics 1993* (Singapore: World Scientific, 1993), pp. 125–136.

1994

'The Measurement Problem,' in L. Accardi (ed.), *The Interpretation of Quantum Theory: Where Do We Stand?* (Rome: Instituto della Enciclopedia Italiana, 1994), pp. 15–24.

'Is Neuropsychology Possible?' in M. Forbes and D. Hull (eds.), *PSA 1994*, Vol. 1, pp. 417–427 (East Lansing: Philosophy of Science Association, 1994).

'Triorthogonal Uniqueness Theorem and its Relevance to the Interpretation of Quantum Mechanics,' *Physical Review A* **49**, 4213–4216 (1994). (With A. Elby.)

'Testing Models of Cognition Through the Analysis of Brain-Damaged Performance,' *British Journal for the Philosophy of Science* **45**, 837–855 (1994).

'On the Structure of Quantal Proposition Systems,' *Foundations of Physics* **24**, 1261–1279 (1994).

'How to Interpret Quantum Mechanics,' *Erkenntnis* **41**, 253–273 (1994).

1995

'Complementarity and the Orthodox (Dirac-von Neumann) Interpretation of Quantum Mechanics,' in R. Clifton (ed.), *Perspectives on Quantum Reality: Non-Relativistic, Relativistic, and Field-Theoretic*, University of Western Ontario Series in Philosophy of Science (Dordrecht: Kluwer, 1995), pp. 211–226.

'Interference, Noncommutativity, and Determinateness in Quantum Mechanics,' *Topoi* **14**, 39–43 (1995).

'Fundamental Problems of Quantum Physics,' *Apeiron* **2**, 98–100 (1995).

'Why Not Take All Observables As Beables?' in *Fundamental Problems in Quantum Theory*, D.M. Greenberger and A. Zeilinger (eds), *Annals of the New York Academy of Sciences* **755**, 761–767 (1995).

'Maximal Structures of Determinate Propositions in Quantum Mechanics,' *International Journal of Theoretical Physics* **34**, 1–10 (1995).

'Quantum Logic,' in R. Audi (ed.), *The Cambridge Dictionary of Philosophy* (Cambridge: Cambridge University Press, 1995), p. 669.

1996

'Modal Interpretations and Bohmian Mechanics,' in J. Cushing, A. Fine, and S. Goldstein (eds.), *Bohmian Mechanics and Quantum Theory: An Appraisal* (Dordrecht: Kluwer, 1996), pp. 331–341.

'A Uniqueness Theorem for Interpretations of Quantum Mechanics,' *Studies in History and Philosophy of Modern Physics* **26** (1996). (With R. Clifton.)

'Schütte's Tautology and the Kochen-Specker Theorem,' *Foundations of Physics* 26, 787–806 (1996).

'Quantum Measurements,' in G.L. Trigg (ed.), *Encyclopedia of Applied Physics,* (VCH Publishers, New York, in collaboration with the American Society of Physics, the German Society of Physics, the Japan Society of Applied Physics, and the Physical Society of Japan, 1996); pp. 257–273. (With D. Greenberger.)

1997

'Schrodinger's Cat and Other Entanglements of Quantum Mechanics,' in J. Earman and J. Norton (eds.), *The Cosmos of Science*, University of Pittsburgh Series in Philosophy of Science (Pittsburgh: University of Pittsburgh Press, 1997), pp. 274–298.

1998

Interpreting the Quantum World (Cambridge: Cambridge University Press, 1997). Winner of the Lakatos Award, 1998.

'The Bare Theory Has No Clothes,' in G. Hellman and R. Healey (eds.), *Quantum Measurement: Beyond Paradox*, Minnesota Studies in Philosophy of Science Vol. XVII (Minneapolis: University of Minnesota Press, 1998), pp. 32–51. (Senior author with R. Clifton and B. Monton.)

'Decoherence in Bohmian Modal Interpretations,' in D. Dieks and P. Vermaas (eds.), *The Modal Interpretation of Quantum Mechanics* (Dordrecht: Kluwer, 1998), pp. 241–252.

'Quantum Measurement Problem,' in E. Craig (ed.), *Encyclopedia of Philosophy* (London: Routledge, 1998).

1999

Interpreting the Quantum World, revised paperback edition (Cambridge: Cambridge University Press, 1999).

Review of *Appearance and Reality: An Introduction to the Philosophy of Physics*, Peter Kosso; *American Journal of Physics*, 1999.

2000

'Quantum Mechanics as a Principle Theory,' *Studies in the History and Philosophy of Modern Physics* **31**.

'Revised Proof of the Uniqueness Theorem for "No Collapse" Interpretations of Quantum Mechanics,' *Studies in History and Philosophy of Modern Physics* **31**, 95–98 (2000). (With R. Clifton and S. Goldstein.)

'Indeterminacy and Entanglement: The Challenge of Quantum Mechanics,' *British Journal for the Philosophy of Science* **51**, 597–615, 2000.

Review of *Quantum Dialogue: The Making of a Revolution*, Mara Beller; *Endeavour* **24**, 179–180 (2000).

2001

'Von Neumann's Theory of Quantum Measurement,' in M. Redei and M. Stötlzner (eds.), *John von Neumann and the Foundations of Quantum Physics*, (Dordrecht: Kluwer, 2001), pp. 63–74.

'Secure Key Distribution via Pre- and Post-selected Quantum States,' *Physical Review A* **63**, 032309–032311 (2001).

'The Quantum Bit Commitment Theorem,' *Foundations of Physics*, **31**, 735–756 (2001).

'Maxwell's Demon and the Thermodynamics of Computation,' *Studies in History and Philosophy of Modern Physics* **32**, 569– 579 (2001).

2002

'Quantum Entanglement and Information,' *The Stanford Encyclopedia of Philosophy* (Winter 2002 Edition), Edward N. Zalta (ed.), URL = http://plato.stanford.edu/archives/win2002/entries/qt-entangle/.

2003

'Indeterminacy and Entanglement: The Challenge of Quantum Mechanics,' in Peter Clark and Katherine Hawley (eds.), *Philosophy of Science Today* (Oxford: Oxford University Press, 2003). (This is a republication of an article with the same title first published in *British Journal for the Philosophy of Science* **51**, 597–615, 2000.)

'Characterizing Quantum Theory in Terms of Information-Theoretic Constraints, *Foundations of Physics* **33**, 1561–1591 (2003). (With Rob Clifton and H. Halvorson.)

Introduction to special issue of *Studies in History and Philosophy of Modern Physics* on quantum information and computation (guest edited by J. Bub and Chris Fuchs), **34B**, 339–342 (2003). (With Chris Fuchs.)

2004

'Why the Quantum?' *Studies in History and Philosophy of Modern Physics* **35**, 241–266 (2004).

Introduction to Special Issue of *Studies in History and Philosophy of Modern* Physics in honor of Rob Clifton, **35B**, 143–149 (2004).

2005

'Quantum Entanglement and Information,' *The Stanford Encyclopedia of Philosophy* (Winter 2006 Edition), Edward N. Zalta (ed.), URL = http://plato.stanford.edu/archives/win2002/entries/qt-entangle/.

'Quantum Mechanics is About Quantum Information,' *Foundations of Physics* **35**, 541–560, 2005.

'Can Cryptography Imply Quantum Mechanics? Reply to Smolin.' *Quantum Information and Computation* **5**, 170–175 (2005). (With H. Halvorson.)

2006

'Quantum Information and Quantum Computing,' forthcoming in John Earman and Jeremy Butterfield (eds.), *Handbook of Philosophy of Physics* (North-Holland, 2006), 103 pp.

'Local Realism and Conditional Probability,' *Foundations of Physics*, 1966. (With Allen Stairs.)

'Copenhagen Interpretation,' in D. Borchert (ed.), *Encyclopedia of Philosophy*, 2nd edition, (Detroit: McMillan Reference USA, 2006).

'Quantum Computing and Teleportation,' in D. Borchert (ed.), *Encyclopedia of Philosophy*, 2nd edition, (Detroit: McMillan Reference USA, 2006).

INDEX